A First Look At IPv6 Hypergiant Infrastructure

Fahad Hilal, Patrick Sattler, Kevin Vermeulen, Oliver Gasser

A First Look At IPv6 Hypergiant Infrastructure

FAHAD HILAL, Max Planck Institute for Informatics, Germany
PATRICK SATTLER, Technical University of Munich (TUM), Germany
KEVIN VERMEULEN, LAAS-CNRS, Université de Toulouse, CNRS, Toulouse, France
OLIVER GASSER, IPinfo, United States and Max Planck Institute for Informatics, Germany

Today's Internet is dominated by a small number of companies which are responsible for a large fraction of Internet traffic. These so called "hypergiants" make use of off-nets to deploy parts of their infrastructure in ISP networks. Off-nets ensure that clients from these ISPs get lower latencies and the ISP needs to send less traffic to its upstream providers. They have been relatively well studied in the IPv4 Internet, although their footprint in IPv6 remains unclear.

In this paper, we take a first look at the IPv6 hypergiant infrastructure. We perform a first-of-its-kind study of IPv6 off-nets for 14 hypergiants and compare their deployment to IPv4. We find IPv6 off-nets in 2k ASes, compared to the more than 6k off-net ASes for IPv4. Moreover, the majority of IPv6 off-nets deployments are seen in ASes which already deploy IPv4 off-nets. Interestingly, we also see some hypergiants such as Disney and Hulu not making use of any IPv6 off-nets at all. We also uncover the phenomenon of cross-hypergiant deployments, where one hypergiant deploys its infrastructure in another hypergiant's network. Finally, we use latency measurements to compare IPv6 vs. IPv4 latency to off-net prefixes within off-net ASes and find similar results for both protocol versions. We make all our code and data available to encourage replicability.

CCS Concepts: • **Networks** → **Network measurement**.

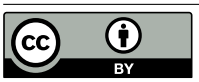
Additional Key Words and Phrases: CDN infrastructure, IPv6, Internet measurement

ACM Reference Format:
Fahad Hilal, Patrick Sattler, Kevin Vermeulen, and Oliver Gasser. 2024. A First Look At IPv6 Hypergiant Infrastructure. *Proc. ACM Netw.* 2, CoNEXT2, Article 11 (June 2024), 25 pages. https://doi.org/10.1145/3656300

1 INTRODUCTION

Delivering content in the Internet has become more and more complex over the years. Whereas in the first days of the Web, websites would simply be hosted in a single location, the advent of Content Delivery Networks (CDNs) have changed this paradigm [55, 56, 67]. These hypergiants leverage deployments of globally distributed servers to deliver content to users from servers close to them. These server deployments can be categorized as being part of the CDN network—i.e., on-net servers—or being deployed in other networks—i.e., off-net servers. Off-net deployments can be seen as a win-win situation for the CDN company as well as the ISP where the off-net servers are deployed. For the ISP, the deployment of off-net servers results in lower latency and thus better quality of experience for their customer, in addition to reducing transit costs. For the CDN, off-nets can be a source of revenue and can lead to reduced bandwidth requirements for on-net servers [32].

Authors' addresses: Fahad Hilal, Max Planck Institute for Informatics, Germany, fhilal@mpi-inf.mpg.de; Patrick Sattler, Technical University of Munich (TUM), Germany, sattler@net.in.tum.de; Kevin Vermeulen, LAAS-CNRS, Université de Toulouse, CNRS, Toulouse, France, kevin.vermeulen@laas.fr; Oliver Gasser, IPinfo, United States and Max Planck Institute for Informatics, Germany, oliver@ipinfo.io.



ACM 2834-5509/2024/6-ART11
https://doi.org/10.1145/3656300

Although the off-net deployment for IPv4 has been studied by Gigis et al. [29], several open questions remain, which we want to address in this work. First, the off-net deployment of IPv6 has not yet been analyzed. As IPv6 is gaining more and more traction [8], this could substantially change off-net deployment overall. Second, off-net deployments have been exclusively studied independently, i.e., cross-hypergiants deployments (e.g., Amazon being deployed at Akamai infrastructure) remain unstudied. Third, it is unclear if there is a performance benefit (or even penalty) when using IPv6 off-nets compared to IPv4 off-nets.

Our main contributions can thus be summarized as follows:

- **IPv6 off-net deployment:** We analyze the deployment of IPv6 off-nets through a first-of-its-kind measurement study. We identify 155k off-net IPv6 addresses in 2k off-net ASes spanning 14 hypergiants. More than 95% of IPv6 off-net ASes deploy IPv4 off-nets as well.
- **Cross-hypergiant deployments:** We identify several cases where hypergiants themselves make use of other hypergiants by deploying infrastructure in their networks. Compared to IPv4, this practice is relatively uncommon in IPv6, with one rare example being Netflix making use of Amazon AWS infrastructure.
- **Performance analysis:** We compare the performance of IPv6 with IPv4 off-net IP addresses and find that top hypergiants exhibit similar latency with the overwhelming part of the differences between IPv4 and IPv6 being under 5 ms.
- **Code and data sharing:** We make all our code and data publicly available, including the list of onnet and offnet IP addresses for IPv4 and IPv6[1].

2 BACKGROUND AND RELATED WORK

Off-nets: Gigis et al. [29] provided a definition for off-nets and conducted a comprehensive analysis of these networks within hypergiants over a span of seven years. In the context of their study, off-nets refer to deployments outside of the hypergiant's own network (e.g., in ISP networks). Their analysis focused specifically on IPv4, utilizing longitudinal data to provide insights into the dynamics of off-net deployments by hypergiants. Gigis et al. collected TLS certificate data and HTTP header information from on-net targets as part of their data acquisition. These datasets were then used to identify off-net hypergiant deployments. We use a similar approach but adapt it where needed for IPv6 measurements.

From another perspective, Vermeulen et al. [64] studied potential issues related to most ISPs colocating off-nets from multiple hypergiants, that could lead to cascading failures. Their study was also exclusively made in IPv4, and replicating their study in IPv6 is out of scope of this paper. Finally, Hasan et al. [32] formalized the tradeoffs between the cost of the cache deployment versus the performance gain with generated topologies.

IPv6 hitlist: Full address space scans as with IPv4 are infeasible on IPv6 due to its large address space. Thus, previous work [26, 27, 50, 70] focused on creating hitlists for IPv6 research. In this work, we utilize the publicly accessible IPv6 hitlist [26]. It includes addresses collected from various sources to build a diverse dataset. Among others, Zirngibl et al. [70] found that fully responsive prefixes (previously called aliased prefixes [26]) also included CDNs prefixes. We use their findings to complement the hitlist with sample addresses picked from the fully responsive prefix dataset.

EDNS0 Client Subnet (ECS): RFC7871 [18] specifies how client subnets can be included in DNS queries. It defines an extension for EDNS0 where the resolver can add the IPv4 or IPv6 client subnet. Resolvers supporting ECS must cache the answers and only use them for queries from these specific subnets. ECS supporting name servers can indicate for which prefix length their answer is valid for (scope prefix length).

[1] *https://doi.org/10.17617/3.64V3MF*

Streibelt et al. [60] and Calder et al. [15] both developed an ECS scanning approach to uncover the infrastructure of Google, while in a more recent study Sattler et al. [52] used ECS to map out the infrastructure of Apple's Private Relay system. Similar to these studies, we use ZDNS [35] to send DNS queries with ECS to complement the data provided by the IPv6 hitlist.

Performance: Several studies have compared the performance of accessing content over IPv4 and IPv6 [5, 10, 12, 43]. Others have also identified service deployments of top players such as Google, Akamai, and Netflix in ISPs. For instance, Bajpai and Schönwälder [11] study the TCP connection establishment time over IPv4 and IPv6 to top 100 popular dual-stacked websites (from Amazon Alexa 1M) from vantage points located in a handful of ASes. In addition to finding the IPv6 performance to have improved, they also observe some of these to be served from CDN caches in ISPs. They also reveal that such caches are largely absent for IPv6. In this work, we find thousands of ISPs deploying IPv6 off-nets from different hypergiants. Doan et al. [20] evaluate the performance of ISP-hosted content caches for Netflix over both address families, finding that such deployments lower the TCP connect times by over 60%. A similar study looks at the latency from dual-stack Sam Knows probes across 74 ASes in traceroutes to GCC nodes streaming YouTube content [21]. While we also perform traceroute-based latency measurements, we carry them out towards thousands of IPv4 and IPv6 off-net deployments across several hypergiants, hosted in hundreds of ISPs.

3 DATASETS AND METHODOLOGY

Prior work has given a clear methodology to uncovering IPv4 on-nets and off-nets, based on TLS and HTTP(s) scans, and we replicate this methodology to uncover the IPv6 on-nets and off-nets. However, the main challenge with IPv6 is that we cannot perform an exhaustive scan of all the IP address space, like in IPv4, and therefore one must rely on publicly available datasets, such as the IPv6 hitlist [26]. Notice that IPv6 hitlist has been designed to reveal responsive IP addresses, and not for our use case. Therefore, in addition to replicating and slightly updating the TLS and HTTP(s) scanning methodology of prior work (Section 3.1), we also perform additional measurements based on DNS and ECS to complete the picture of on-net and off-net deployments (Section 3.2).

3.1 Adapting IPv4 TLS and HTTP(S) Scans to IPv6

Gigis et al. [29] proposed a two step approach to unveil the IPv4 on-nets and off-nets. The first step consists of collecting fingerprints from on-net IP addresses of the hypergiants. The second step consists of comparing these on-net fingerprints with fingerprints from IP addresses outside hypergiant ASes to detect off-net IP addresses. We detail about how these next steps work and highlight the updates that we apply to prior methodology in the following paragraphs. As in prior work [29], we focus on the following hypergiants: Akamai, Alibaba, Amazon, Apple, Cdnetworks, Chinacache, Disney, Fastly, Google, Hulu, Incapsula, Limelight, Meta, Microsoft, Netflix, Twitter, and Yahoo. In our private discussions with Cloudflare, we learnt that Cloudflare does not deploy off-nets at the time of our measurements[2] so we exclude it from our investigation.

Replicating IPv4 Fingerprint Collection Methodology: Gigis et al. [29] collect TLS and HTTP(S) fingerprints from on-net prefixes. A prefix is considered an on-net prefix if it maps to the hypergiant organization. They use data from RouteViews and RIPE RIS to map prefixes to ASes, and CAIDA's AS2Org dataset [41] to map these ASes to organizations.

To establish the on-net TLS fingerprint of a hypergiant, we first find the certificates containing the hypergiant's name in the `Organization` field of the `Subject Name`, and then extract the list of all DNS names of the certificate from the subject alternative name field. The TLS fingerprint is the union of the different DNS names found in the different certificates of a hypergiant.

[2]We also learnt they plan to deploy offnets in the future.

To establish the HTTP(S) fingerprints of a hypergiant, we extract the HTTP(S) headers specific to a hypergiant found on the on-net IP addresses. As in prior work, we remove common standard headers (e.g., `Cache-Control`) and inspect the remaining headers from our scans to establish per-hypergiant fingerprints. We identify headers that occur frequently across more than 50 on-net IP addresses and ensure that they are used exclusively by a specific hypergiant to consider it as a fingerprint. Like in prior work, we then perform manual analysis and pick those that either have an abbreviated name or value of the hypergiant or are documented or disclosed online.

Updating Fingerprint Collection Methodology: We find that certificates for Amazon and Google do not have a value for the "Subject Organization" field. To address this, we use the certificates which have "google" or "amazon" in their DNS names field covered by the certificate. Furthermore, we also collect 70 new HTTP(S) headers compared to prior work. This is expected as these headers are likely to change over time [29]. We find official documentation for over 60% of these headers.

Adapting Fingerprint Collection Methodology to IPv6: Instead of using Rapid7 [48] and Censys [22] datasets to collect the TLS and HTTP(S) scans like prior work [29], we perform our own large-scale TLS and HTTP(S) scans, both for IPv4 and IPv6. This choice is motivated by the fact that recent work has shown that Rapid7 and Censys datasets are not as reliable for IPv6 as they are in IPv4 [7, 51]. We scan the entire address space for IPv4, as done in prior work, and we scan the publicly available IPv6 hitlist [26, 59, 70] for IPv6. We use the list of all known IPv6 addresses for the latter and also scan fully responsive prefixes which are frequently seen in hypergiants [53, 70]. We first perform ZMap [23] scans to identify responsive IP addresses on port 80 and 443 and then use ZGrab2 [61] to fetch TLS certificates and collect the HTTP(S) headers.

To complement the set of IP addresses found by the IPv6 hitlist scan, we perform additional DNS measurements and ECS-enabled DNS measurements described in Section 3.2. We know that these IP addresses are by definition used to serve a hypergiant's service, so if the resolved IP address belongs to the hypergiant, we consider it an on-net, whereas we consider it as an off-net if the IP addresses belongs to another organization. We then merge the TLS and HTTP(S) on-net fingerprints from the IPv6 hitlist scans with the ones from the DNS-based measurements. The goal is to improve the TLS and HTTP(S) fingerprints as the IPv6 hitlist is not designed to unveil the hypergiant's infrastructure. All measurements were performed in November, 2023.

Replicating Off-net Detection IPv4 Methodology: If an IP address matches a hypergiant TLS fingerprint and contains at least one HTTP(S) header from the fingerprints, and the IP address is not an on-net IP address, it is classified as an off-net IP address of that hypergiant.

3.2 Improving the IPv6 Coverage With DNS Measurements

We find indications that the IPv6 hitlist should have a good coverage of the hypergiant's infrastructure as it is designed to find responsive IP addresses, and many hypergiant prefixes are fully responsive [69]. To confirm this, we perform additional regular and ECS-enabled DNS measurements to enrich the set of on-net TLS and HTTP(S) fingerprints that serve to find IPv6 off-nets, and eventually to reveal IPv6 off-nets themselves. First, our regular DNS measurements consist of resolving the top 10k Google CrUX [17] domains using MassDNS [14], followed by extracting the TLS certificates and the HTTP(S) headers from the resolved addresses, through ZGrab2 from our vantage point at the Max Planck Institute in Germany. We retrieve the CrUX toplist from July 2023. Second, we leverage the EDNS Client Subnet (ECS) option which allows to specify the IP prefix of the client in a DNS query. This option helps service providers to tailor the responses of their authoritative server based on the prefix of the client, rather than based on the recursive resolver. This is particularly useful when the client uses a public resolver, which can be geographically far away from the client. Prior work demonstrated that some hypergiantss use ECS to redirect their clients, at least in IPv4 [15, 68], so it can reveal the hypergiants' infrastructure [15]. Our

ECS-enabled DNS measurements uses the same idea as prior work [15], i.e., varying the ECS prefix in the DNS queries to find different IP addresses used by the hypergiants. However, running these IPv6 ECS-enabled DNS measurements presents some challenges, which we describe next.

Prefix Choice for ECS Queries: In IPv4, one can perform an exhaustive scan by using all the possible /24 prefixes in the ECS measurements. In IPv6, this is not possible, so we have to choose which prefixes to use for our ECS queries. We use a combination of three datasets studied in prior work [34]: BGP prefixes, bulk WHOIS data, and the /56 prefixes of the IPv6 addresses found in the IPv6 hitlist. The three datasets are likely complementary as they do not provide the same view of the IPv6 landscape: BGP prefixes provide a list of routed prefixes, while the bulk WHOIS data provides a list of IPv6 prefix allocation to their organization [34], and usually contains more specific prefixes than the routed ones. Finally, the /56 prefixes from the hitlist generally contain the most specific prefixes, but might not contain all the routed space as it only reveals responsive IP addresses. Finding the perfect set of prefixes for ECS measurements in IPv6 is out of scope, as the purpose of our measurements is to map the hypergiants' IPv6 infrastructure, and so our findings should be interpreted as a lower bound of what can be unveiled with ECS.

We collect BGP prefixes from a RIB snapshot from RouteViews [1] on November 17, 2023, while the bulk WHOIS data and the IPv6 hitlist are from October and September 2023. The BGP snapshot contains 203,629 prefixes, and among them 203,336 prefixes have a length of 56 or less, which is the most specific prefix length that is recommended to be put in ECS queries according to RFC 7871 [18]. The bulk WHOIS data from October 13, 2023, containing data from all five RIRs (RIPE, APNIC, ARIN, LACNIC, and AFRINIC) and RADB, has 1,589,014 prefixes, with 1,002,080 prefixes with a length of 56 or less. The IPv6 hitlist contains 9,533,649 responsive IP addresses, from which we extract 1,704,435 unique /56 prefixes. We send ECS queries for the top 60 domain names of the Google CrUX toplist [17] that we identify as belonging to the hypergiants passing the TLS and HTTP(S) fingerprints. These 60 domain names belong to 8 hypergiants, i.e., Akamai, Amazon, Apple, Fastly, Google, Meta, Microsoft, and Netflix. We run the ECS measurements using the Google Public DNS resolver, as it is known to forward ECS queries to the authoritative servers and is whitelisted [6]. The queries are sent at a rate of 1,000 queries per second using ZDNS [35]. We run all ECS measurements in November 2023.

Results: For each IP address revealed by the ECS measurements, we look at whether this IP address is an on-net address of a hypergiant, using the methodology described in Section 3.1. If this is the case, we look at the set of DNS names covered by the TLS certificate, and the set of the HTTP(S) headers of the IP address. With the DNS measurements, we find no additional DNS names and a total of three additional HTTP(S) headers.

These fingerprints gives us only five more off-net IP addresses which we do not have in off-net addresses from the hitlist. These results show that the TLS and HTTP(S) fingerprints already present through measurements with the IPv6 hitlist are sufficient to have a good coverage of the IPv6 on-net infrastructure of the hypergiants. We discuss more details of our ECS results in Section 5.

3.3 Ethics and Reproducibility

Before we conduct our measurements, we incorporate proposals by Partridge and Allman [44] and Kenneally and Dittrich [36]. We follow best measurement practices [23] by limiting our probing rate, using a well-established blocklist, and making use of dedicated measurement servers. These measurement servers communicate the scientific nature of our measurements with an informing rDNS name, a website providing more information, and contact details to reach out to us in case of issues. During our measurements, we received a handful of mostly automated abuse emails, to which we replied promptly.

We plan to publish our measurement results along with the analysis scripts to foster reproducibility in networking research.

4 IPV6 OFF-NET DEPLOYMENT

In this section, we present our results of our Internet-scale IPv6 off-net deployment study and contrast it to the IPv4 off-net deployment. We report these footprints in terms of the number of IP addresses as well as the ASes. Furthermore, we study the continent coverage of these footprints. We also investigate the types of ASes that different hypergiants choose to deploy their off-net edge servers in. In addition to analyzing the footprint from a networking and geographical stand point, we also shed light on the Internet user-base that has access to these off-net deployments. See Table 6 in Appendix B for an overview of our IPv6 and IPv4 TLS scans. Even if the analysis is done at the IP address level, it only gives us a binary indication of the presence of an off-net in an AS or not. It does not provide a quantitative metric for the deployment, as one IP address could be mapped to multiple servers. For this reason, our analysis is performed at AS level.

4.1 Hypergiant Off-net Infrastructure

The IPv4 and IPv6 off-net IP addresses and ASes we uncover across different hypergiants based on TLS only as well as TLS and header validations are detailed in Tables 4 and 5 in Appendix A. In our analysis, we only use the more strict latter validation unless otherwise explicitly stated. We refer to such off-net deployments throughout the text as off-nets or off-nets with server installations (as per Gigis et al. [29]).

Off-net Deployment: Table 1 shows the number of off-net ASes and off-net IP addresses for different hypergiants. In total, we find 155k IPv6 and 357k IPv4 off-net addresses in 2k unique IPv6 and 6k IPv4 off-net ASes. Moreover, we find Google, Netflix, and Meta to be the only hypergiants which deploy off-nets in more than 1000 ASes. In fact, the Google off-net footprint in terms of the number of ASes is nearly double that of the other two. The remaining hypergiants deploy off-nets in a handful (e.g., Twitter) or hundreds (e.g., Microsoft, Apple) ASes.

We find the most IPv6 off-net addresses for Amazon. In fact, Amazon's IPv6 off-net addresses are close to four times higher than its IPv4 off-net count (68.7k vs. 18.1k), making it the only hypergiant with more IPv6 off-net addresses than IPv4 addresses. The vast majority (51.1k out of 68.7k) of Amazon IPv6 off-net addresses are found in AS 174 (Cogent). We investigate if this large number of Amazon off-net IP addresses is substantially influenced by fully responsive prefixes, as prior work showed that Amazon had a large number of fully responsive prefixes [26]. Of the 1.7M scanned prefixes, there is only a single Amazon prefix containing off-net IP addresses, constituting 0.4% of the total number of off-net IP addresses. Looking at other HGs, only a total of seven fully responsive prefixes have off-net IP addresses, for a total of 545 IP addresses. These results show that our analysis is not biased by fully responsive prefixes. Other hypergiants such as Google, Meta, Akamai, and Alibaba, have a substantially lower number of IPv6 off-net addresses compared to their IPv4 deployments. For Netflix, however, the drop from the IPv4 numbers is not that steep. Finally, we also find smaller number (under 500) of IPv6 off-net addresses for Apple, Microsoft, and Fastly.

At an AS-level, we observe Amazon's over 65k IPv6 off-net addresses to be deployed in only 11 ASes, compared to its 171 IPv4 off-net ASes. Google and Meta both deploy IPv6 off-nets in more than 1k ASes, with Netflix slightly below 1k. Other notable IPv6 deployments are found by Akamai in 241 and Apple in 117 ASes. The remaining hypergiants deploy IPv6 off-nets in less than 40 ASes. Compared to IPv4, the number of IPv6 off-net ASes is lower for all observed hypergiants. Moreover, a large portion of IPv6 off-net ASes also hosts IPv4 off-net prefixes, i.e., IPv6 ASes are to a large degree a subset of IPv4 off-net ASes.

	IP addresses		ASes		
Hypergiant	IPv6	IPv4	IPv6	IPv4	Both
Akamai	24274	121813	241	881	223
Alibaba	7711	68217	37	175	26
Amazon	68674	18116	11	171	7
Apple	396	1600	117	219	104
Disney	0	39	0	2	0
Fastly	246	27	2	6	0
Google	24738	64370	1342	4976	1291
Hulu	0	13	0	1	0
Meta	16017	65942	1231	2565	1185
Microsoft	49	521	2	174	0
Netflix	5625	11569	928	2731	860
Twitter	2	6	2	5	1
Union	155161	357535	2043	6004	1959

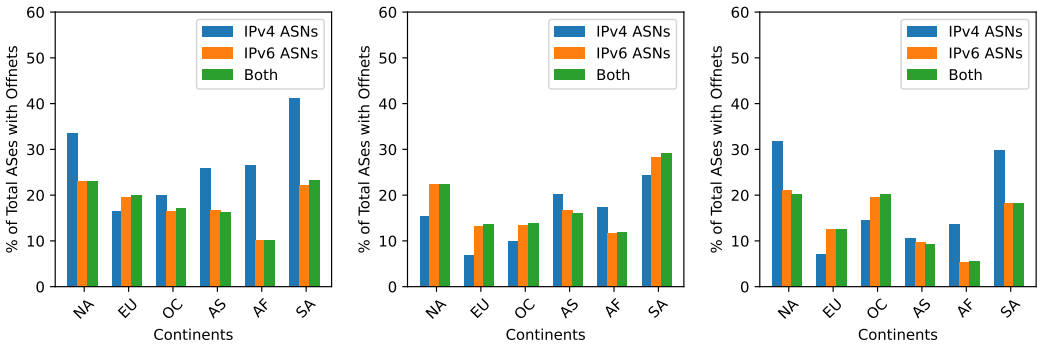
Table 1. Number of IPv6 and IPv4 off-net addresses and ASes (TLS + Header validated) per hypergiant.

Comparing the number of our off-net IPv4 ASes to previous work [29] shows that the off-net footprint for pretty much all hypergiants has been growing at a decent pace. Among the top three, Google and Netflix both show a growth of about 30% from 2021 to 2023, while Meta only increases by half that percentage. Apple which had no off-net ASes in 2021 now has over 200 while Amazon shows an increase in the off-net IPv4 AS footprint by 175%. However, even though Akamai has a nearly 900 off-net ASes, the decrease in its off-net footprint revealed by Gigis et al. [29] still continues which is evident from a 19.5% drop.

Takeaway: *Off-nets are deployed in a lower number of IPv6 ASes compared to IPv4 ASes across all hypergiants. Majority of hypergiants deploy IPv6 off-nets in ASes where IPv4 off-nets already exist.*

Geographical Off-net Footprint: Next, we map off-net ASes to the continents that they serve. Firstly, we use the off-net addresses and map them to their ASes while also geolocating them to a particular country. We use the MaxMind GeoIP2 Country database [40] for the latter. While IP geolocation is known to be somewhat inaccurate when it comes to city-level or even more fine-granular geolocation [28, 31, 47, 54], we think it adds valuable information to country and continent-level off-net deployments. We then treat the country as the country served by the AS and then map it to its continent. This implies that the same AS can be seen across different continents which is common observation in the current Internet [62, 66]. We focus our per-continent analysis on off-nets by Google, Netflix, and Meta—the top three based on the number of ASes deploying their off-nets. For all three, most of the off-net ASes are located in South America, with around 500 IPv6 and more than 1000 IPv4 off-net ASes. The number is particularly striking for Google which deploys IPv4 off-nets in over 1.8k ASes in the region. Oceania is found to have the least number of deployed IPv6 off-net ASes for all top-three hypergiants. See Figures 9a to 9c in Appendix C for details on the absolute number of off-net ASes.

Figures 1a to 1c show the fraction that Google, Meta, and Netflix off-net ASes make of the total number of ASes serving different continents. We use APNIC's ASpop dataset [37, 38] to find all ASes (IPv4 and IPv6 separately) that serve a non-zero user base across different countries belonging to the six continents. Contrary to the absolute numbers, the relative analysis shows that North America has the highest fraction of IPv6 off-net covered ASes for Google and Netflix with around 20% of all IPv6 eyeball ASes. Meta has the highest IPv6 AS coverage in South America with more than 30%, followed by North America with slightly less than 25% of all eyeball ASes. The remaining



(a) Google. (b) Meta. (c) Netflix.

Fig. 1. Fraction of ASes that serve the region and also deploy off-nets for the top three hypergiants.

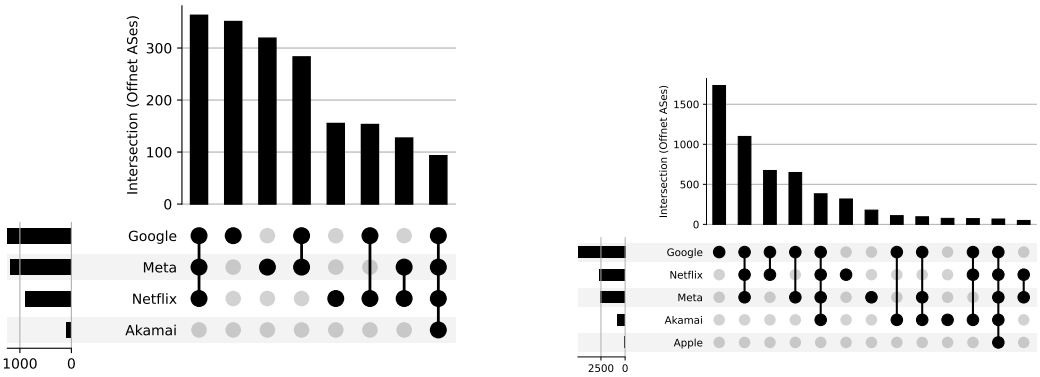
continents are covered by about 10%–20% by off-nets. For all top-three hypergiants, the worst IPv6 coverage can be seen in Africa with 5%–13%. When comparing the AS continent coverage to IPv4, we find that IPv4 is most of the time outpacing IPv6. This is especially prominent in Africa where the IPv6 coverage is substantially lower, e.g., for Google IPv6 off-nets in Africa are found in around 10% of ASes, compared to almost 30% in IPv4. One exception to IPv4's dominance over IPv6 is Meta which has four continents seeing higher IPv6 coverage. Another observation is that IPv6-only off-net ASes seem to be rare across all regions.

We also enrich our analysis by using CAIDA's AS2Org dataset [41] which maps ASes to their parent organizations which are registered in a single country[3]. The results are similar to the aforementioned findings. Additionally, this shows an expected slight drop in the numbers (as some ASes are merged as siblings), but we observe no change in trends across continents.

We further investigate the aggressive deployment of off-nets in South America and check if this is correlated with low on-net presence on the continent. Thus, we look at the number of on-net and off-net addresses of the top three hypergiants, in South America. The number of off-net IPv6 addresses are always substantially larger than the number of on-nets. For Google, IPv6 on-net addresses contribute 5.3% to all of Google's on-nets, but they are still behind the number of IPv6 off-net deployments in South America which make up nearly 16%. For Meta, we find no IPv6 on-net addresses on the continent whereas the IPv6 off-net addresses contribute over 30% to Meta's global off-net deployments. Finally, Netflix's IPv6 on-net deployments (4.6%) also trail its off-net deployments (29.5%) by a huge margin. A similar picture can be seen for IPv4 also, with even more pronounced differences between the number of on-net and off-net addresses in South America.

Moreover, we compare our findings with external data, i.e., the "CDN and Cloud Infrastructure Location" dataset [58]. The dataset provides data on the global infrastructural deployments of major players like Google, Meta, Akamai, and Netflix. The data is obtained from various sources such as websites of service providers or airport code hints in URLs. For Google and Meta, we are unable to perform the on-net vs. off-net analysis from the dataset as the on-nets can not be distinguished from off-nets from the available data. Additionally, for these the dataset also lacks IPv6 deployment data. However, we find all relevant data for Netflix. We look for the number of Netflix Open Connect Appliances (OCA) in the dataset that are Netflix on-nets or off-nets within the South American continent (over both address families). For IPv6, we find 38 IPv6-capable on-net and 791 off-net

[3]Note that an organization's registered country is not necessarily the same or only one where it has an IP footprint.



(a) IPv6. (b) IPv4.

Fig. 2. Number of Hypergiants hosted by off-net ASes.

AS Type	IPv6	IPv4
Access	1879	4832
Content	39	303
Transit/Access	77	379
Enterprise	28	171
Tier-1	7	9
n/a	13	310

Table 2. Off-net AS classification based on network type.

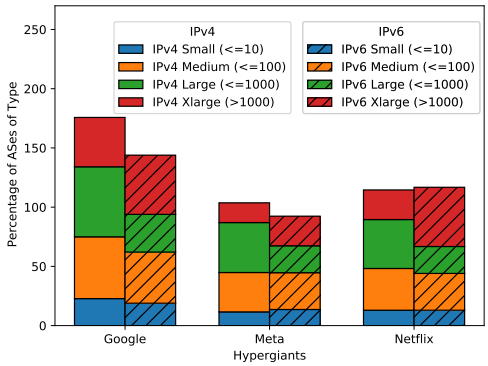


Fig. 3. Top 3 hypergiants and their AS types based on customer cone sizes ASes.

OCAs across South America. The ISPs with these off-net ASes make up more than 50% of all IPv6 off-net ASes. For IPv4, there are 39 on-net and 650 off-net OCAs in South America; more than 90% of these are in Brazil. These ISPs make up over 40% of all ISPs with Netflix OCAs. This confirms our hypothesis that the aggressive deployment of off-nets in South America stems from the relatively poor on-net presence and the less developed peering infrastructure on the continent [25].

While these analyses show the fraction of ASes covered by off-net deployments per region, they do not provide any insights on the Internet population of users served by off-nets in these regions. Although the number of ASes being covered by off-nets for a region might be small, the Internet-user base served can still be substantial if the off-nets are deployed in ASes that cover a larger fraction of users. We discuss this aspect in more detail in Section 4.4.

Takeaway: *We find that top hypergiants (HGs) deploy off-nets globally. The top-three HGs deploy off-nets in more than 18% of all ASes in North and South America. For Meta, IPv6 off-nets see a higher share of AS coverage compared to IPv4 in most continents. The large off-net deployment in South America is correlated with lower on-net deployment there.*

Off-nets from Multiple Hypergiants: We investigate the phenomenon of ASes hosting IPv6 off-nets from multiple hypergiants in more detail and show the results in Figure 2a. We only depict results for the top 4 hypergiants with most off-net ASes to increase the readability. We observe that the most common case is that ASes host off-nets from all top-three hypergiants. Other common non-singleton combinations are two of Google, Meta, and Netflix. Deploying an Akamai off-net alongside these is relatively rare with under 5% of these ASes observed to do so.

In IPv4, we find that of all ASes that host off-nets from the top 6 hypergiants based on the number of off-net ASes, most of them always host off-nets for Google either exclusively or in combination with other hypergiants (see Figure 2b). For instance, 1756 (30.6%) of such 5.6k ASes solely host Google off-nets but then Google off-nets are also paired with Netflix and Meta off-nets (Google, Meta, Netflix) by about 23% (1313) of ASes. The numbers are comparable to the findings of Vermeulen et al. [64] with our numbers being slightly lower due to the application of the header validation as well. However, we observe reluctance in also hosting an off-net for a fifth IPv4 hypergiant from the set of Apple, Alibaba, and Microsoft. In fact we only find 1.5% (87) of all ASes (5.8k), which host an off-net from at least one hypergiant from the top seven, to host off-nets for more than four hypergiants in addition to the top four.

Takeaway: *It is relatively common to have different hypergiants deploy off-nets in the same AS. This is especially the case for Google, Meta, and Netflix for IPv6; for IPv4, Akamai is also more present along with the top-3 in IPv6.*

4.2 Off-net AS Classification

Next, to better understand in which ASes off-nets are deployed, we classify them with regards to the AS type as well as its customer-cone size.

AS Types: In addition to mapping off-net ASes to continents, we also classify them into network types. Table 2 depicts the categories the ASes map to using CAIDA's AS Classification dataset from 2021[4]. Since we find the dataset to offer good coverage for our analysis and ASes are quite unlikely to change their types, we choose to perform our analysis with this dataset nevertheless its age.

We find IPv6 off-nets to be predominantly deployed in access networks (see Table 2). This is quite expected as hypergiants intend to deploy servers close to users. This deployment strategy also suits smaller access networks wishing to cut down on the inter-AS traffic. There are, however, also Tier-1 ASes hosting off-nets. For instance, AS 701 (Verizon UUNET) hosts IPv6 off-nets for Google and Meta, in addition to also hosting IPv4 off-nets. Additionally, Tier-1s like AS 7018 (AT&T) and AS 6461 (Zayo) only host IPv6 off-nets. The former hosts for several hypergiants like Akamai, Amazon, Netflix, Google, Microsoft, Alibaba, Meta, and Apple whereas the latter only has Akamai off-nets.

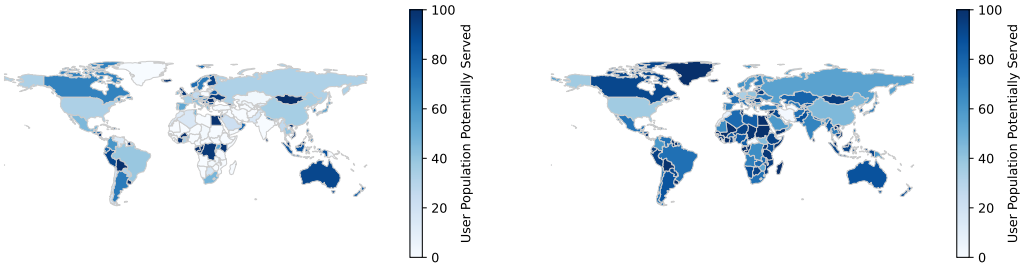
Other larger ASes which are sometimes considered as Tier-1s also host IPv6 off-net IP addresses. For instance, AS 4134 (Chinanet Backbone) and AS 174 (Cogent) deploy a much higher number of IPv6 off-net IP addresses. Overall, we find that all Tier-ones that host both IPv4 and IPv6 off-nets host them for the same hypergiants.

We validate our findings by performing the classification also with datasets from other sources including PeeringDB [46], ASdb [71], and BGP tools [13]. All these datasets agree with with access networks being the most prominent choice for deploying IPv6 and IPv4 off-nets.

Takeaway: *Hypergiants deploy their IPv6 off-nets in all different types of networks, but the vast majority can be found in access networks. A handful of Tier-1s can surprisingly also be found deploying off-nets. The picture in IPv4 looks similar compared to IPv6.*

AS Customer-Cone Size: Motivated by previous works [16, 29], we also classify the IPv6 and IPv4 off-net ASes based on their customer cone sizes. We use the CAIDA's AS relationship dataset [39],

[4]The updates to the dataset were discontinued in 2021.



(a) IPv6. (b) IPv4.

Fig. 4. Google: Fraction of a country's Internet users in ASes hosting off-nets.

which gives us the customer-provider and peer-to-peer AS relationships for IPv6 as well as IPv4. Unlike the AS relationship peer-provider-determined-cone (PPDC) dataset, this dataset does not allow us to see Stub ASes (ASes with no customer cones). However, the PPDC dataset only provides IPv4 AS relationships. Therefore, we decide against using it. As a result, we do not look at stub ASes. We break ASes into 4 types: Small ASes (customer cone $\leq$ 10 ASes), medium ASes (customer cone $\leq$ 100 ASes), large ASes (customer cone $\leq$ 1000 ASes), and extra large ASes (customer cone $>$ 1000 ASes). We show the results in Figure 3.

For IPv6, small ASes dominate with nearly 2.1k (86%) such ASes while medium ASes contribute 13%. Large and extra large ASes make up 0.9% and 0.1%, respectively. For Google, the vast majority of IPv6 off-nets are found in small ASes (73.3 %). However, this could be biased by an uneven distribution of AS types in the dataset. Therefore, we also investigate, how many ASes of each type are covered by Google, and find that medium ASes are best covered with 43% deploying Google's IPv6 off-nets. While extra large ASes still cover half, large and small ASes are only 31.8% and 18% covered, respectively. For Meta, most of the IPv6 off-nets ASes are again small ASes but like Google's IPv6 footprint, medium ASes are most covered (31%). Finally, Netflix's IPv6 off-net deployments in different types of ASes follow Meta's. We only find Netflix to deploy off-nets in one additional extra large AS compared to Meta (two out of four) making it the most covered followed by medium sized ASes at 31%.

For IPv4, we also find small ASes making up the bulk with 84% of the dataset, medium and large AS contribute 13.9% and 1.9%, respectively. Similar as in IPv6, Google deploys the majority of its off-nets in small ASes (69%), followed by medium (26.3%) and large ASes (4.1%). Contrary to IPv6, Google has best IPv4 coverage in large ASes (60%), with around 50% coverage for medium ASes. Similarly, Meta best covers large (42.2%) and medium ASes (33.2%) Finally, Netflix mostly mimics Meta in IPv4—same as in IPv6—in terms of covering different types of ASes with off-net deployments and we again see large ASes (41.3%) being the most covered.

Takeaway: *Small ASes are dominating the hypergiant deployment in IPv6 as well as IPv4. Coverage-wise we see differences: In IPv6 hypergiants best cover small ASes (customer cone $\leq$ 10 ASes), in IPv4 medium (customer cone $\leq$ 100 ASes) and large ASes (customer cone $\leq$ 1000 ASes) are best covered.*

4.3 Cross-Hypergiant Deployments

We would usually expect hypergiant (HG) off-nets to be deployed inside ISPs ASes and not other HG ASes. However, we find this to be the case for several hypergiants, see Table 3. Enhanced reliability of services, traffic re-routing and increased coverage could be some of the factors behind such decisions. We keep the term off-net as it fits the formal definition given in Section 3 of an HG

	Off-net IP addresses/Other HG		Fraction of all Off-net IP addresses	
Hypergiant	IPv4	IPv6	IPv4	IPv6
Akamai	Google (52), Amazon (12)	-	0.06%	-
Alibaba	Amazon (15), Microsoft (6)	Google (1)	0.04%	0.01%
Amazon	Alibaba (22), Microsoft (19)	-	0.3%	-
Apple	Amazon (14), Google (14)	Google (4), Microsoft (1)	2.7%	1.3%
Disney	Amazon (39)	-	100%	-
Fastly	Amazon (6), Microsoft (1)	-	25.9%	-
Google	Amazon (121), Akamai (26)	Akamai (5)	0.03%	0.02%
Hulu	Amazon (13)	-	100%	-
Meta	Google (10), Alibaba (4)	Google (15)	0.03%	0.09%
Microsoft	Amazon (27), Alibaba (10)	-	9%	-
Netflix	Amazon (253), Google (25)	Amazon (42), Google (21)	2.4%	1.1%
Twitter	Akamai (6)	Akamai (1)	33.3%	50%

Table 3. Number of off-nets of a hypergiant (HG) in other hypergiants (only top two) and the percentage such off-net IP addresses present across other hypergiants contribute to total off-net footprint.

hosting a server in another organization. It could also be that the cross hypergiant off-net is used for a side activity and not core business. However, we cannot distinguish between these different cases as it is hard to know which services are served by an off-net (see Section 4.5).

In IPv6, this strategy of hypergiants deploying their off-nets in other hypergiants is relatively uncommon. We only see Netflix with more than 20 off-net addresses in some other hypergiants. Out of the 67 such off-net IP addresses, 42 are hosted in Amazon ASes. While we see no off-net deployments for Hulu and Disney, we observe Microsoft and Apple using Akamai over IPv6.

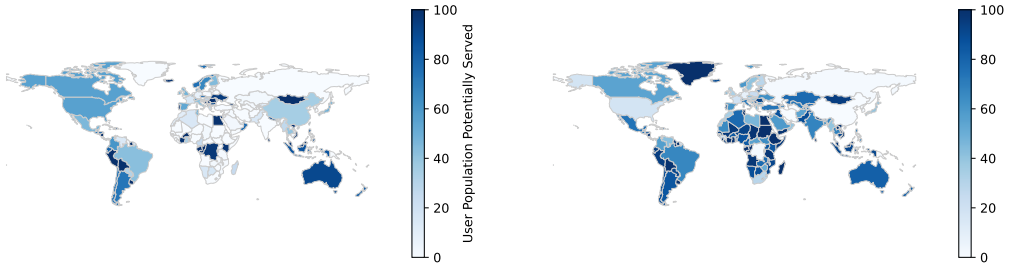
In IPv4, this practice is much more common. For instance, we see Amazon using Alibaba, Akamai, and Google among others. However these only make up less than 1% of its off-net deployment. We investigate the DNS names present in the certificates we receive for these IP addresses and find that they are predominantly responsible for serving the "amazon.com" web shop. Similarly, we observe "apple.com" or "images.apple.com" to be visible in 2.7% of Apple's off-nets deployed on Amazon, Alibaba, Google, and Akamai. For other players such as Hulu and Disney, for which we find under 100 off-nets, we observe all of them to be in Amazon ASes. Looking at the DNS names for the latter, we find the keyword "disney" in all of them, however, the exact purpose of the DNS names is difficult to ascertain (e.g., `origin.prod.mdxpepui.mdx.las1.wdpro.disney`). These findings are expected as these hypergiants have been known to rely upon multiple players including Akamai and Amazon for delivering their video streaming services [19].

For the top most off-net deploying hypergiant, Google, we find 0.3% of its off-net footprint across Akamai, Amazon, Microsoft, Alibaba, and Apple. 121 out of these 196 off-net IP addresses are in Amazon and based on the DNS names appear to be delivering Google services such as Google Analytics, the DoubleClick Ad service, 'google.com', and 'youtube.com'. Finally, we observe 2.4% of Netflix's off-nets (with server deployments) in Amazon, Google, or Microsoft. In fact, over 94% of such IP addresses belong to Amazon, which confirms findings by previous work [4].

Takeaway: *For the first time we characterize cross-HG deployments—i.e., HGs deployments in other HG networks— and find them to be relatively uncommon in IPv6. We see Netflix making use of Amazon AWS infrastructure in IPv6. In IPv4 this practice is much more common, where we see a single-digit percentage share of some HGs being deployed in other HGs.*

4.4 Internet User Population

We also investigate the fraction of Internet population per country that has access through IPv4 and IPv6 to HG services hosted in their network provider. We again use APNIC's ASpop dataset [37, 38]



(a) IPv6. (b) IPv4.

Fig. 5. Meta: Fraction of a country's Internet users in ASes hosting off-nets.

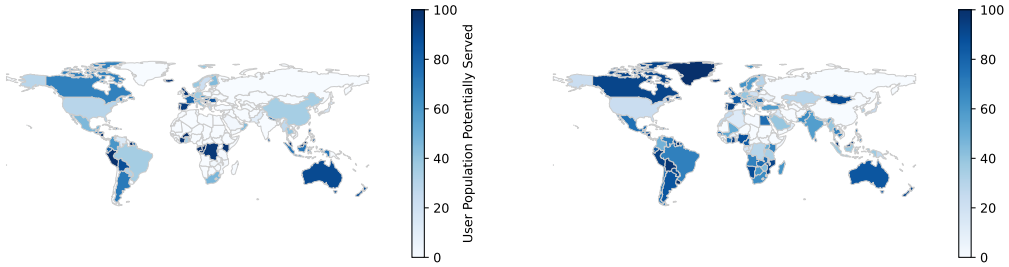
which gives the market share of users served by an AS per country. For our analysis, we sum the market share of all ASes that host hypergiant off-nets (TLS + HTTP(S) validated) and operate in the country. We focus on Google, Netflix, and Meta which we find to deploy off-nets in most ASes.

For Google, Figure 4a show good IPv6 coverage across all continents, with Africa being the least well covered. In IPv4 (cf. Figure 4b), the coverage is much more dense compared to IPv6. Overall, the IPv6 coverage seems to be catching up instead of complementing the IPv4 coverage. The only exception is the Democratic Republic of the Congo (DRC) where IPv6 coverage far exceeds IPv4 coverage (96.2% vs. 62.4%). This does not seem to be a conscious effort to favor IPv6; instead all three ASes that host IPv6 off-nets also host IPv4 off-nets but their IPv4 share of the user population over IPv6 is much larger. On a continental level, while IPv6 user population coverage is exemplary in Oceania and Europe, and above average in the Americas, it can be substantially improved in Asia and Africa. For instance, deploying IPv6 off-nets in AS 7922 (Comcast) and AS 51659 (Baxet) can already enhance the coverage by over 25 percentage points in the US and by about 20 percentage points in Russia, respectively. The only major deviation from the 2021 results is an increase of nearly 20% of Google's IPv4 coverage in China.

The user population coverage by Meta's IPv6 off-nets almost mimics that of Google's IPv6 deployment (except Canada -12%) as is depicted in Figure 5a. In other similarities to Google, the DRC case also appears for Meta. In contrast to Google, Russia is largely underserved by Meta's IPv6 off-nets. In fact, we notice Meta off-nets in only two Russian ASes—AS 3216 (Vimpelcom), hosting IPv4 and IPv6, and AS 49070 (Rostelecom) which only has IPv4 off-nets. These have a collective market share of less than 1% in both IPv6 and IPv4. More measured placement of off-nets could greatly enhance the user population coverage. For instance, deploying IPv6 off-nets in AS 51659 (Baxet) and AS 8359 (MTS) can improve IPv6 coverage in Russia by over 50% and IPv4 off-nets in AS 57354 (Systema) and AS 50257 (A-Mobile) by nearly 25%. Like Google, Meta's IPv4 coverage (see Figure 5b) is largely unaffected from the 2021 results except for a slight increase in China.

For Netflix (see Figure 6a), the user population that has access to Netflix services through IPv6 off-nets lags behind Google but is not far off that of Meta. Similarly to Google and Meta, we continue to see the catching-up effect in IPv6 for Netflix. On the IPv4 side (see Figure 6b), we see Netflix to consolidate its footing in most of the countries it was observed to deploy its off-nets. This is particularly true in Canada, Australia,and several South American countries (e.g., Argentina, Peru, Bolivia) where the IPv4 user base now exceeds 80% per country.

For all top-three hypergiant IPv6 off-net deployments, one common observation is that India appears to be under-served. This is interesting because India has a rapidly growing IPv6 deployment [57]. Closer investigation reveals that there are 48 ASes that operate in India and also host



(a) IPv6. (b) IPv4.

Fig. 6. Netflix: Fraction of a country's Internet users in ASes hosting off-nets.

IPv6 off-nets of different HGs. We find a total of 2,149 off-net IPv6 addresses of which nearly 56% are in AS 9498 (Bharti Airtel). We find all of these ASes to have a a very small market share in comparison to ASes such as AS 55836 (Reliance Jio) and AS 45609 (Bharti Airtel GPRS). Placing IPv6 off-nets in these ASes can boost the user base by over 85%.

Looking in more detail into the deployment in China, we analyze whether censorship could play a role in off-net deployment. For instance, it is known that China imposes censorship on Facebook and Google [24, 33, 42, 45, 63]. We compare the fraction of ASes with off-nets and users according to the APNIC ASpop dataset between countries known to impose censorship and countries deemed to be censorship-free. For instance, just over 1% and 3% of the ASes which have a non-zero user base in China deploy Meta's IPv4 and IPv6 off-nets respectively, with 7% and 3% of the ASes for Google, respectively. Iran, also previously found to censor Meta [9, 65], simply has no Meta off-nets, and 8% of IPv4 and 0% of IPv6 ASes in Iran deploy Google off-nets. On the other hand, countries less likely to impose censorship such as the US have close to 9% and 11% of such ASes deploying Meta's IPv4 and IPv6 off-nets, respectively. This effect is even more pronounced for Google where 41% and 18% of the ASes deploy Google's IPv4 and IPv6 off-nets, respectively. We also looked at some European countries where there are no known instances of censorship (Sweden, Finland, Denmark, Norway), and found an average of 6% IPv4 and 10% IPv6 ASes with off-nets for Meta, and 22% IPv4 and 21% IPv6 ASes with off-nets from Google. Although the numbers differ between countries that are known to impose censorship and those that are deemed to be censorship-free, the lower deployment in censored countries could also be due to business decisions of HGs.
Takeaway: *The top three HGs have IPv6 deployments covering users in different countries reasonably well. In comparison to IPv4, however, IPv6 has still some catching up to do, especially in countries in Africa and Asia. We identify a small number of ASes, where off-net deployment can have a large impact by reaching millions of users.*

4.5 Services Served by Off-nets

To better understand the use of off-nets, we investigate whether we can measure which services are served by an off-net. For each off-net IP address, we extract DNS names from the TLS certificate. If this set only contains a single name and is not a wildcard, then we know that this off-net serves this particular domain and service.

For IPv4, of the 357,535 off-net addresses, we find that 8,869 (2.5%) have a single non-wildcard domain name in the certificate, and for IPv6, of the 155,161 IPv6 off-net addresses, there are 609 (0.39%) of them. First, this shows that it is hard to identify which services are run by off-nets using TLS certificates. Second, in discussions with a French ISP having Akamai and Google off-nets, we

learnt that they were not aware of which clients or services were served by the off-nets, showing that—most likely—only the HGs know the full picture.

For the off-net IP addresses where we can identify the service, we find 1,915 IPv4 off-net IP addresses that serve `google.com`, which is interesting because prior work stated that `google.com` is not served by off-nets anymore [64]. For Netflix, we find that `ichnaea-web.netflix.com` is the most frequent domain name, which appears to be a tracker to collect client information[5]. For Meta, we have no IP addresses where we can identify the service. For IPv6, the same happens for Google and Meta. For Netflix, the most frequent hostname is the same as in IPv4.

Takeaway: *From the TLS certificates, it is hard to infer which services are served by off-nets, especially in IPv6, where only 0.39% of the off-net IP addresses have an identifiable service.*

4.6 Off-nets and ROV

Prior work has shown that the off-nets of different HGs can be colocated, representing an increased risk of spillover over peering and transit links in case of failure [64]. Along those lines of increased risk, we look at the security aspect, which is another type of risk. As for operational expertise and resources that differ between ISPs and HGs, ISPs might also take more time to adopt ROV to protect their prefixes. However, HGs could require ISPs to perform ROV on the prefixes used for the off-nets. Although we do not find any requirement to have ROV on an off-net prefix in the documentation of the off-nets for Google and Netflix [2, 30], for each AS hosting an off-net, we compare two fractions (Table 7 in Appendix D): the number of BGP prefixes covered by ROV over the total number of BGP prefixes (ROV BGP), and the number of off-net BGP prefixes covered by ROV over the total number of off-net BGP prefixes (BGP Off-net ROV) of this AS for which we find at least one off-net IP address. We compute these two fractions both for IPv4 and IPv6. We retrieve the ROV data of December 1, 2023 using rpki-validator [3].

For the big three HGs (Google, Meta, and Netflix), we find that Meta and Netflix for IPv4 have more ASes where the fraction of off-net BGP prefixes with ROV is larger than the overall fraction of BGP prefixes covered by ROV, with 52% and 58%. For all of the other pairs of (HG, IP version), it is the opposite, with up to 70% of the ASes in IPv6 for Meta having a smaller fraction of off-net prefixes covered by ROV.

Takeaway: *There is no evidence that Google, Meta, or Netflix have requirements for protecting off-net prefixes with ROV, potentially representing a security risk.*

5 RESULTS FROM ECS MEASUREMENTS

In Section 3.2, we stated that our ECS measurements did not bring much additional coverage to the IPv6 hitlist. In this section, we provide more details about the ECS measurement results: In short, we find that our three datasets of ECS prefixes are complementary to perform ECS measurements, and demonstrate that the ECS measurements did not return off-net IP addresses because almost all the domain names that we used were not served by off-nets. More details about the ECS behaviors of different hypergiants can be found in Appendix E.

For the eight hypergiants that we probe with ECS, Netflix and Fastly both respond to our DNS ECS queries with a scope of 0, meaning that the answer returned by their authoritative name server does not depend on the ECS prefix we sent. As these two HGs are not ECS-enabled, we do not consider them for the remainder of the ECS analysis.

How Complementary Are Our Set of Prefixes? Figure 7 shows Venn diagrams of the overlap between the sets of IP addresses and BGP prefixes found by the three datasets used as input for

[5]We infer the meaning of the domain name by finding user PiHole reports on the Web and noticing that "ichnaea" means tracker in Greek.

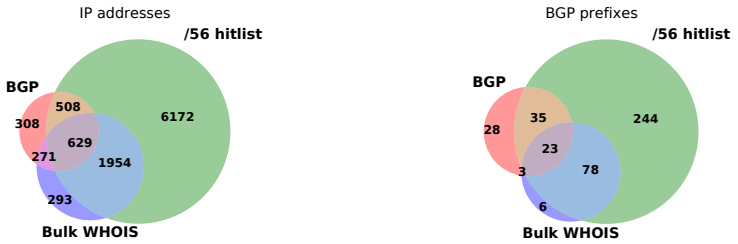


Fig. 7. Venn diagrams showing the complementarity of our three datasets of prefixes with different metrics: IP addresses (left), and BGP prefixes (right).

the ECS prefixes. We choose to not only look at the set of IP addresses but also look at a higher granularity, the BGP prefixes, because some hypergiants (e.g., Amazon) can return random IP addresses within a prefix [26, 70]. As a result, if we only look at the number of IP addresses, it could be possible that one of the datasets of prefixes used for ECS brings more discoveries just because it is larger. Aggregating the addresses into their BGP prefixes reduces this bias. To differentiate the resulting BGP prefixes from the ones in the ECS queries, we call the latter the ECS BGP prefixes.

We see that the /56 prefixes from the hitlist bring more discoveries in general, with 92% of the BGP prefixes that are seen. The two other datasets also bring some unique discoveries. An interesting finding is that despite the set of ECS BGP prefixes being 5 times smaller than the bulk WHOIS prefixes, it brings 24 unique BGP prefixes versus 7. These numbers represent 6% and 2% of the total number of BGP prefixes revealed by the three datasets.

Takeaway: *Our three ECS prefix datasets are complementary to reveal the IPv6 infrastructure.*

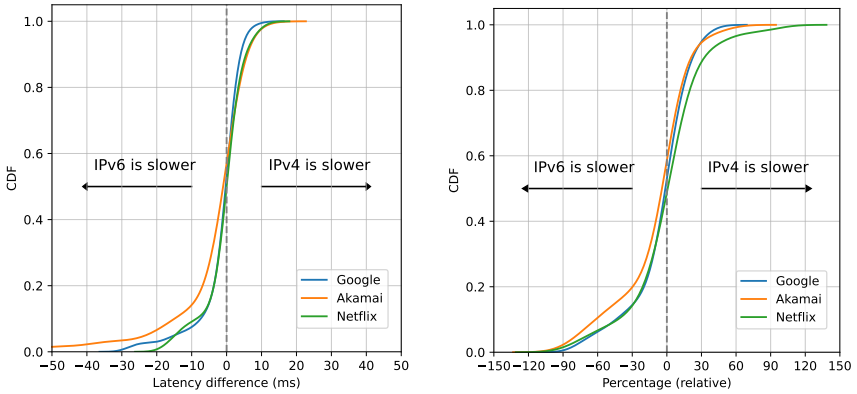
Are The Chosen Domain Names Served by Off-nets? An authoritative server can respond to our ECS query with an on-net or an off-net IP address. For all the IP addresses found in our ECS measurements, we look at their TLS and HTTP(S) fingerprints to see whether the authoritative server returned an off-net or an on-net IP address. Of the 10k IP addresses revealed by our ECS measurements, only 35 are off-net IP addresses, with 13 from Apple and 22 from Akamai, and correspond to two domain names: “swcdn.apple.com” and “www.akamai.com”. To confirm that our ECS queries do not have a special treatment compared to queries from local resolvers, we run queries to the same domain names from the ≈ 2k IPv6 capable RIPE Atlas probes belonging to ASes with off-nets using their local resolvers. We find similar results as with our ECS queries, with only four off-net IP addresses revealed, two for Apple and two for Akamai.

Takeaway: *Almost all the 60 popular base domain names belonging to HGs are not served by off-nets.*

How Well Suited is the IPv6 Hitlist to Find Off-nets? Given that we use IPv6 addresses from the IPv6 hitlist [26] to discover off-nets, we also investigate the suitability of the hitlist for this task. As the previous section showed that we cannot reveal off-nets with ECS measurements, we instead look at how well suited the hitlist is compared to ECS to find *on-nets*. We find that 75% of on-net BGP prefixes from our ECS measurements are also present in the IPv6 hitlist. This percentage is likely even higher for off-net prefixes, as the hitlist is known to perform better for ISP infrastructure (e.g., off-nets) compared to large CDN operators (e.g., on-nets) [26]. We leave further investigation to find better target lists for IPv6 off-net discovery for future work.

6 HYPERGIANT OFF-NET PERFORMANCE

In addition to uncovering the off-net deployments in IPv6 and comparing them to IPv4, we also evaluate and compare their performance through latency-based measurements. An investigation of



(a) Latency difference ($IPv4 - IPv6$) for different hypergiants.

(b) Relative Latency ($\frac{IPv4-IPv6}{max(IPv4,IPv6)}$) difference for different hypergiants.

Fig. 8. Comparison of IPv4 and IPv6 latencies seen in off-net ASes for different hypergiants.

this nature can reveal if IPv6 off-net deployments are at par with IPv4 ones, are catching up or if there remains significant room for improvement. Previous studies [20, 21] have focused on singular hypergiants like Netflix or Google to study the performance of cache deployments within ISPs while also using a smaller number of vantage points. Although we discuss the top hypergiants with the largest IPv4 and IPv6 off-net footprints, our methodology generalizes to all the hypergiants we study. Additionally, our probes are more numerous and diverse in terms of their ASes with there being nearly 500 dual-stacked probes across around 100 ASes. Furthermore, in private communications with several hypergiants, they confirmed that the content served over both address families is the same. Thus, this is the first study analyzing IPv6 vs. IPv4 performance of off-nets at scale.

Methodology: We start by identifying the ASes that host dual-stacked off-nets of the hypergiants under analysis. Subsequently, we conduct a search for dual-stack RIPE Atlas probes within these identified ASes. Utilizing the aforementioned dual-stack probes, we then proceed to perform traceroute measurements specific to each hypergiant. These measurements encompass both IPv4 and IPv6 off-net targets situated within the AS of the probes. Our rationale for adopting this methodology is grounded in the observations presented in Section 4.1, where we establish that hypergiants tend to deploy off-nets close to users within eyeball networks.

The traceroute measurements include up to 25 off-net IPv4 and IPv6 addresses per AS and are targeted towards TCP/443. We include at most ten RIPE Atlas probes per available AS, and we select the probes based on their reliability and latency. We rank the probes based on the sum of their average latencies (IPv4 and IPv6) to the first two public hops as seen in "built-in measurement" [49] traceroutes towards k-root DNS servers.

Results: For our analysis, for each AS with successful traceroutes, we average latencies towards all IPv4 off-nets (average IPv4 latency) and all IPv6 off-nets (average IPv6 latency). To minimize the impact of outliers, we only consider off-nets which have more than one probe successfully completing a traceroute measurement. We consider a measurement to be successful if the probe receives a SYN-ACK from the off-net. Moreover, we only consider ASes which have such successful traceroutes to at least five IPv4 and IPv6 off-net addresses within the same AS.

Out of the nearly 1.9k ASes which host IPv4 and IPv6 off-nets, we find dual-stack RIPE Atlas probes in a total of 585 ASes (30%). We get 437 ASes for Google, 304 for Netflix, 313 for Meta,

and 182 for Akamai (the top four hypergiants by largest dual-stack off-net footprint). However, we do not see successful traceroutes to IPv4 and IPv6 off-nets for each AS even though we use dual-stacked probes. This is not unusual as the RIPE Atlas probes availability can vary over time. Owing to this, we are left with 218 (49.9%) Google, 156 (51.4%) Netflix, and 109 (59.9%) Akamai off-net ASes where at least one IPv4 and IPv6 target could be reached by multiple probes. Meta presents a curious case as we find that its IPv4 and IPv6 off-nets rarely respond to our probes. In fact, even when repeating the measurement, we only see one such relevant AS for Meta.

Following the application of our filtering criteria (i.e., at least five successful IPv4 and IPv6 traceroutes per AS), we are left with 73 (-66.5%) off-net ASes for Google, 60 (-68%) for Akamai, 43 (-60%) for Netflix, and none for Meta. Although the remaining set of ASes is considerably smaller, we choose to err on the side of caution and only analyze ASes that fulfill our filtering requirements.

We plot the absolute and relative difference in latencies (IPv4 – IPv6) for Google, Netflix, and Akamai in Figure 8. For both Google and Akamai, the latency to IPv4 off-nets is lower in most ASes (57.5% and 58.3%, respectively). However, the average IPv4 latency is lower than the average IPv6 latency by only 5 ms for close to three-fourth of these ASes for both hypergiants. Additionally, IPv4 is faster by 10 ms or fewer in nearly 90% of ASes for Google whereas for Akamai this is only the case for about 77% of ASes. Finally, no ASes have an inferior IPv6 latency by more than 30 ms for either hypergiant. Nearly 40% of the ASes for both hypergiants, have a better latency for IPv6 off-nets. For Google for over 95% of the cases the latency is lower by about 5ms, whereas for Akamai this is the case for a slightly lower 88% of the ASes.

For Netflix, the percentage of off-net ASes with a better average IPv6 latency (53.4%) although higher is not far from the ASes which have a better average IPv4 latency (46.5%). For ASes of the former class, the latency is again better by no more than 5 ms in nearly 85% of the cases. Additionally, the average IPv6 latency is better by between 5 to 20 ms in just over 15% of the case. For ASes with a better average IPv4 latency, exactly three-fourth have a better latency again only by under 5 ms. In fact, the average IPv6 latency never lags by more than 20 ms.

Finally, we validate our results if we only keep the ASes where the distribution of the latency over the different probes is close to the mean. Indeed, for some ASes, the average latency is not representative of the latency experienced by a client, typically when the standard deviation is significant compared to the mean. We perform the same analysis as above, removing ASes where the standard deviation is more than p times the mean RTT, p varying from 0.1 to 0.9. The lower p is, the fewer ASes we keep in our dataset. The number of ASes that we keep for these different values of p varies from 73 down to 17 for Google, 60 to 15 for Akamai, and 43 to 5 for Netflix. For Google, the percentage of ASes where IPv4 is better than IPv6 varies from 57.5% to 70.6%, while it is 53.3% to 60% for Akamai, and 38.5% to 60% for Netflix. Although the numbers vary significantly for Netflix, the vast majority of the ASes have a small latency difference, with 84-100% of the ASes having a latency difference under 5 ms for Google, 80-97% for Akamai, and 79-92% for Netflix

Takeaway: *Top hypergiants exhibit similar IPv6 as IPv4 performance for their off-nets, with the overwhelming majority of latency differences being under 5 ms.*

7 CONCLUSION

In this paper, we took a first look at the IPv6 off-net infrastructure of 14 hypergiants. We found 155k IPv6 off-net addresses in more than 2k ASes. Moreover, the majority of IPv6 off-nets deployments were seen in ASes which already deployed IPv4 off-nets. We also uncovered the phenomenon of cross-hypergiant deployments, where one hypergiant deploys its infrastructure in another hypergiant's network. Finally, we used latency measurements and found similar latencies when connecting to an off-net address over IPv6 compared to IPv4.

Acknowledgements. We thank the anonymous reviewers and our shepherd for their valuable feedback. This work was partially funded by the German Federal Ministry of Education and Research under project PRIMEnet (16KIS1370). We also thank the CNRS and the Max-Planck-Gesellschaft for funding Fahad to visit the LAAS-CNRS via the SALTO exchange program

REFERENCES

[1] [n. d.]. Oregon Route Views. http://routeviews.org/.
[2] 2024. Introduction to GGC. https://support.google.com/interconnect/answer/9058809.
[3] 2024. RPKI validator. https://rpki-validator.ripe.net.
[4] Vijay Kumar Adhikari, Yang Guo, Fang Hao, Matteo Varvello, Volker Hilt, Moritz Steiner, and Zhi-Li Zhang. 2012. Unreeling netflix: Understanding and improving multi-cdn movie delivery. In *2012 Proceedings IEEE Infocom*. IEEE, 1620–1628.
[5] Saba Ahsan, Vaibhav Bajpai, Jörg Ott, and Jürgen Schönwälder. 2015. Measuring YouTube from dual-stacked hosts. In *Passive and Active Measurement: 16th International Conference, PAM 2015, New York, NY, USA, March 19-20, 2015, Proceedings 16*. Springer, 249–261.
[6] Rami Al-Dalky, Michael Rabinovich, and Kyle Schomp. 2019. A Look at the ECS Behavior of DNS Resolvers. In *Proceedings of the Internet Measurement Conference*. 116–129.
[7] Taha Albakour, Oliver Gasser, and Georgios Smaragdakis. 2023. Pushing Alias Resolution to the Limit. In *ACM Internet Measurement Conference 2023*. https://doi.org/10.1145/3618257.3624840
[8] APNIC. 2023. *IPv6 BGP table*. https://bgp.potaroo.net/v6/as2.0/index.html
[9] Simurgh Aryan, Homa Aryan, and J Alex Halderman. 2013. Internet censorship in Iran: A first look. In *3rd USENIX Workshop on Free and Open Communications on the Internet (FOCI 13)*.
[10] Vaibhav Bajpai, Saba Ahsan, Jürgen Schönwälder, and Jörg Ott. 2017. Measuring YouTube content delivery over IPv6. *ACM SIGCOMM Computer Communication Review* 47, 5 (2017), 2–11.
[11] Vaibhav Bajpai and Jürgen Schönwälder. 2015. IPv4 versus IPv6-who connects faster?. In *2015 IFIP Networking Conference (IFIP Networking)*. IEEE, 1–9.
[12] Vaibhav Bajpai and Jürgen Schönwälder. 2019. A longitudinal view of dual-stacked websites—failures, latency and happy eyeballs. *IEEE/ACM Transactions on Networking* 27, 2 (2019), 577–590.
[13] bgp.tools. 2023. *bgp.tools*. https://bgp.tools/
[14] Blechschmidt, S. 2024. *MassDNS*. https://github.com/blechschmidt/massdns
[15] Matt Calder, Xun Fan, Zi Hu, Ethan Katz-Bassett, John Heidemann, and Ramesh Govindan. 2013. Mapping the Expansion of Google's Serving Infrastructure. In *Proceedings of the 2013 Conference on Internet Measurement Conference* (Barcelona, Spain) *(IMC '13)*. Association for Computing Machinery, New York, NY, USA, 313–326.
[16] Matt Calder, Xun Fan, Zi Hu, Ethan Katz-Bassett, John Heidemann, and Ramesh Govindan. 2013. Mapping the expansion of Google's serving infrastructure. In *Proceedings of the 2013 conference on Internet measurement conference*. 313–326.
[17] Chrome User Experience Report contributors. 2023. *Chrome User Experience Report*. https://developer.chrome.com/docs/crux/
[18] Carlo Contavalli, Wilmer van der Gaast, David C Lawrence, and Warren "Ace" Kumari. 2016. Client Subnet in DNS Queries. RFC 7871. https://doi.org/10.17487/RFC7871
[19] Investor's Business Daily. 2023. *Disney Stock, Akamai Stock: Streaming Video Key For Both*. https://www.investors.com/news/technology/disney-stock-akamai-stock-video-streaming/
[20] Trinh Viet Doan, Vaibhav Bajpai, and Sam Crawford. 2020. A longitudinal view of Netflix: Content delivery over IPv6 and content cache deployments. In *IEEE INFOCOM 2020-IEEE Conference on Computer Communications*. IEEE, 1073–1082.
[21] Trinh Viet Doan, Ljubica Pajevic, Vaibhav Bajpai, and Jorg Ott. 2018. Tracing the path to youtube: A quantification of path lengths and latencies toward content caches. *IEEE Communications Magazine* 57, 1 (2018), 80–86.
[22] Z. Durumeric, D. Adrian, A. Mirian, M.Bailey, and J. A. Halderman. 2015. A Search Engine Backed by Internet-Wide Scanning. In *ACM Conference on Computer and Communications Security (CCS)*.
[23] Zakir Durumeric, Eric Wustrow, and J Alex Halderman. 2013. ZMap: Fast Internet-wide Scanning and Its Security Applications. In *22nd USENIX Security Symposium (USENIX Security 13)*. 605–620.
[24] Oliver Farnan, Alexander Darer, and Joss Wright. 2016. Poisoning the well: Exploring the great firewall's poisoned dns responses. In *Proceedings of the 2016 ACM on Workshop on Privacy in the Electronic Society*. 95–98.
[25] Hernan Galperin. 2013. Connectivity in Latin America and the Caribbean: The role of internet exchange points. (2013).
[26] Oliver Gasser, Quirin Scheitle, Pawel Foremski, Qasim Lone, Maciej Korczyński, Stephen D Strowes, Luuk Hendriks, and Georg Carle. 2018. Clusters in the Expanse: Understanding and Unbiasing IPv6 Hitlists. In *Proceedings of the*

Internet Measurement Conference 2018. 364–378.
[27] Oliver Gasser, Quirin Scheitle, Sebastian Gebhard, and Georg Carle. 2016. Scanning the IPv6 Internet: Towards a Comprehensive Hitlist. In *Traffic Monitoring and Analysis Workshop 2016*.
[28] Manaf Gharaibeh, Anant Shah, Bradley Huffaker, Han Zhang, Roya Ensafi, and Christos Papadopoulos. 2017. A look at router geolocation in public and commercial databases. In *Proceedings of the 2017 Internet Measurement Conference*. 463–469.
[29] Petros Gigis, Matt Calder, Lefteris Manassakis, George Nomikos, Vasileios Kotronis, Xenofontas Dimitropoulos, Ethan Katz-Bassett, and Georgios Smaragdakis. 2021. Seven years in the life of Hypergiants' off-nets. In *Proceedings of the 2021 ACM SIGCOMM 2021 Conference*. 516–533.
[30] Google. 2024. Requirements for deploying embedded appliances. https://support.google.com/interconnect/answer/9058809.
[31] Bamba Gueye, Steve Uhlig, and Serge Fdida. 2007. Investigating the imprecision of IP block-based geolocation. In *Passive and Active Network Measurement: 8th Internatinoal Conference, PAM 2007, Louvain-la-neuve, Belgium, April 5-6, 2007. Proceedings 8*. Springer, 237–240.
[32] Syed Hasan, Sergey Gorinsky, Constantine Dovrolis, and Ramesh K Sitaraman. 2014. Trade-offs in optimizing the cache deployments of CDNs. In *IEEE INFOCOM 2014-IEEE conference on computer communications*. IEEE, 460–468.
[33] Nguyen Phong Hoang, Arian Akhavan Niaki, Jakub Dalek, Jeffrey Knockel, Pellaeon Lin, Bill Marczak, Masashi Crete-Nishihata, Phillipa Gill, and Michalis Polychronakis. 2021. How Great is the Great Firewall? Measuring China's {DNS} Censorship. In *30th USENIX Security Symposium (USENIX Security 21)*. 3381–3398.
[34] Amanda Hsu, Frank Li, and Paul Pearce. 2023. Fiat lux: Illuminating IPv6 apportionment with different datasets. *Proceedings of the ACM on Measurement and Analysis of Computing Systems* 7, 1 (2023), 1–24.
[35] Liz Izhikevich, Gautam Akiwate, Briana Berger, Spencer Drakontaidis, Anna Ascheman, Paul Pearce, David Adrian, and Zakir Durumeric. 2022. ZDNS: A Fast DNS Toolkit for Internet Measurement. In *Proceedings of the 22nd ACM Internet Measurement Conference* (Nice, France) *(IMC '22)*. Association for Computing Machinery, New York, NY, USA, 33–43.
[36] Erin Kenneally and David Dittrich. 2012. The Menlo Report: Ethical Principles Guiding Information and Communication Technology Research. *Available at SSRN 2445102* (2012).
[37] APNIC Labs. 2023. *APNIC Labs AS Popularity Statistics*. https://stats.labs.apnic.net/aspop/
[38] APNIC Labs. 2023. *APNIC Labs IPv6 Popularity Statistics*. https://stats.labs.apnic.net/v6pop/
[39] M Luckie, B Huffaker, k claffy, A Dhamdhere, and V Giotsas. 2013. AS Relationships, Customer Cones, and Validation. In *ACM Internet Measurement Conference (IMC)*. 243–256. https://doi.org/10.1145/2504730.2504735
[40] MaxMind. 2023. *MaxMind GeoIP Database Documentation*. https://dev.maxmind.com/geoip/docs/databases/city-and-country
[41] Netflix. 2024. CAIDA AS2Org dataset. https://openconnect.zendesk.com/hc/en-us/articles/360034538352-Requirements-for-deploying-embedded-appliances.
[42] Arian Akhavan Niaki, Nguyen Phong Hoang, Phillipa Gill, Amir Houmansadr, et al. 2020. Triplet Censors: Demystifying Great {Firewall's}{DNS} Censorship Behavior. In *10th USENIX Workshop on Free and Open Communications on the Internet (FOCI 20)*.
[43] Mehdi Nikkhah, Roch Guérin, Yiu Lee, and Richard Woundy. 2011. Assessing IPv6 through web access a measurement study and its findings. In *Proceedings of the Seventh conference on emerging Networking EXperiments and Technologies*. 1–12.
[44] Craig Partridge and Mark Allman. 2016. Ethical considerations in network measurement papers. *Commun. ACM* 59, 10 (2016), 58–64.
[45] Paul Pearce, Ben Jones, Frank Li, Roya Ensafi, Nick Feamster, Nick Weaver, and Vern Paxson. 2017. Global measurement of {DNS} manipulation. In *26th USENIX Security Symposium (USENIX Security 17)*. 307–323.
[46] PeeringDB. 2023. *PeeringDB*. https://www.peeringdb.com/
[47] Ingmar Poese, Steve Uhlig, Mohamed Ali Kaafar, Benoit Donnet, and Bamba Gueye. 2011. IP geolocation databases: Unreliable? *ACM SIGCOMM Computer Communication Review* 41, 2 (2011), 53–56.
[48] Rapid7. 2023. *Open Data*. https://opendata.rapid7.com/
[49] RIPE NCC. 2023. *RIPE Atlas Built-in Measurements Documentation*. https://atlas.ripe.net/docs/built-in-measurements/
[50] Erik Rye and Dave Levin. 2023. IPv6 Hitlists at Scale: Be Careful What You Wish For. In *Proceedings of the ACM SIGCOMM 2023 Conference* (New York, NY, USA) *(ACM SIGCOMM '23)*. Association for Computing Machinery, New York, NY, USA, 904–916.
[51] Said Jawad Saidi, Srdjan Matic, Oliver Gasser, Georgios Smaragdakis, and Anja Feldmann. 2022. Deep Dive into the IoT Backend Ecosystem. In *ACM Internet Measurement Conference 2022*. https://doi.org/10.1145/3517745.3561431
[52] Patrick Sattler, Juliane Aulbach, Johannes Zirngibl, and Georg Carle. 2022. Towards a Tectonic Traffic Shift? Investigating Apple's New Relay Network. In *Proceedings of the 22nd ACM Internet Measurement Conference* (Nice, France) *(IMC '22)*.

Association for Computing Machinery, New York, NY, USA, 449–457.
[53] Patrick Sattler, Johannes Zirngibl, Mattijs Jonker, Oliver Gasser, Georg Carle, and Ralph Holz. 2023. Packed to the Brim: Investigating the Impact of Highly Responsive Prefixes on Internet-wide Measurement Campaigns. In *ACM Conference on emerging Networking EXperiments and Technologies 2023*. https://doi.org/10.1145/3629146
[54] Quirin Scheitle, Oliver Gasser, Patrick Sattler, and Georg Carle. 2017. HLOC: Hints-based Geolocation Leveraging Multiple Measurement Frameworks. In *Network Traffic Measurement and Analysis Conference 2017*.
[55] Brandon Schlinker, Hyojeong Kim, Timothy Cui, Ethan Katz-Bassett, Harsha V Madhyastha, Italo Cunha, James Quinn, Saif Hasan, Petr Lapukhov, and Hongyi Zeng. 2017. Engineering egress with edge fabric: Steering oceans of content to the world. In *Proceedings of the Conference of the ACM Special Interest Group on Data Communication*. 418–431.
[56] Kyle Schomp, Onkar Bhardwaj, Eymen Kurdoglu, Mashooq Muhaimen, and Ramesh K Sitaraman. 2020. Akamai dns: Providing authoritative answers to the world's queries. In *Proceedings of the Annual conference of the ACM Special Interest Group on Data Communication on the applications, technologies, architectures, and protocols for computer communication*. 465–478.
[57] Internet Society. 2023. *Four of the World's Top 10 Populous Countries Driving IPv6 Adoption*. https://pulse.internetsociety.org/blog/four-of-the-worlds-top-10-populous-countries-driving-ipv6-adoption
[58] Steve Song. 2023. Cloud CDN Cache. https://github.com/stevesong/cloud_cdn_cache.
[59] Lion Steger, Liming Kuang, Johannes Zirngibl, Georg Carle, and Oliver Gasser. 2023. Target Acquired? Evaluating Target Generation Algorithms for IPv6. In *Proceedings of the Network Traffic Measurement and Analysis Conference (TMA)* (Naples, Italy).
[60] Florian Streibelt, Jan Böttger, Nikolaos Chatzis, Georgios Smaragdakis, and Anja Feldmann. 2013. Exploring EDNS-Client-Subnet Adopters in Your Free Time. In *Proceedings of the 2013 Conference on Internet Measurement Conference* (Barcelona, Spain) *(IMC '13)*. Association for Computing Machinery, New York, NY, USA, 305–312.
[61] The ZMap Project. 2023. *ZGrab 2.0*. https://github.com/zmap/zgrab2
[62] Mehmet Engin Tozal. 2016. The Internet: A system of interconnected autonomous systems. In *2016 Annual IEEE Systems Conference (SysCon)*. IEEE, 1–8.
[63] Elisa Tsai, Deepak Kumar, Ram Sundara Raman, Gavin Li, Yael Eiger, and Roya Ensafi. 2023. CERTainty: Detecting DNS Manipulation using TLS Certificates. In *Privacy Enhancing Technologies Symposium (PETS)*.
[64] Kevin Vermeulen, Loqman Salamatian, Sang Hoon Kim, Matt Calder, and Ethan Katz-Bassett. 2023. The Central Problem with Distributed Content: Common CDN Deployments Centralize Traffic In A Risky Way. In *Proceedings of the 22nd ACM Workshop on Hot Topics in Networks*. 70–78.
[65] Matthäus Wander, Christopher Boelmann, Lorenz Schwittmann, and Torben Weis. 2014. Measurement of globally visible dns injection. *IEEE Access* 2 (2014), 526–536.
[66] Uri Yacobi-Keller, Evgeny Savin, Benjamin Fabian, and Tatiana Ermakova. 2019. Towards Geographical Analysis of the Autonomous System Network. *International Journal of Networking and Virtual Organisations* 21, 3 (2019), 379–397.
[67] Kok-Kiong Yap, Murtaza Motiwala, Jeremy Rahe, Steve Padgett, Matthew Holliman, Gary Baldus, Marcus Hines, Taeeun Kim, Ashok Narayanan, Ankur Jain, et al. 2017. Taking the edge off with espresso: Scale, reliability and programmability for global internet peering. In *Proceedings of the Conference of the ACM Special Interest Group on Data Communication*. 432–445.
[68] Jiangchen Zhu, Kevin Vermeulen, Italo Cunha, Ethan Katz-Bassett, and Matt Calder. 2022. The best of both worlds: high availability CDN routing without compromising control. In *Proceedings of the 22nd ACM Internet Measurement Conference*. 655–663.
[69] Johannes Zirngibl, Philippe Buschmann, Patrick Sattler, Benedikt Jaeger, Juliane Aulbach, and Georg Carle. 2021. It's Over 9000: Analyzing Early QUIC Deployments with the Standardization on the Horizon. In *Proceedings of the 21st ACM Internet Measurement Conference*. 261–275.
[70] Johannes Zirngibl, Lion Steger, Patrick Sattler, Oliver Gasser, and Georg Carle. 2022. Rusty Clusters? Dusting an IPv6 Research Foundation. In *Proceedings of the 22nd ACM Internet Measurement Conference* (Nice, France) *(IMC '22)*. Association for Computing Machinery, New York, NY, USA, 395–409.
[71] Maya Ziv, Liz Izhikevich, Kimberly Ruth, Katherine Izhikevich, and Zakir Durumeric. 2021. ASdb: a system for classifying owners of autonomous systems. In *Proceedings of the 21st ACM Internet Measurement Conference*. 703–719.

Hypergiant	IPv4 (TLS, TLS + HTTP(S))	IPv6 (TLS, TLS + HTTP(S))
Akamai	(122931, 121813)	(24332, 24274)
Alibaba	(74441, 68217)	(10307, 7711)
Amazon	(59763, 18116)	(68952, 68674)
Apple	(64811, 1600)	(856, 396)
Disney	(23692, 39)	(61, 0)
Cdnetworks	(316, 0)	(3, 0)
Fastly	(53, 27)	(246, 246)
Google	(67234, 64370)	(27757, 27438)
Hulu	(1930, 13)	(2, 0)
Meta	(74225, 65942)	(16205, 16017)
Microsoft	(79988, 521)	(1333, 49)
Netflix	(18877, 11569)	(6830, 5625)
Twitter	(2420, 6)	(3, 2)
Yahoo	(76, 0)	(3, 0)

Table 4. Number of off-net IPs (TLS validated, TLS + header validated) per hypergiant.

Hypergiant	IPv4 (TLS, TLS + HTTP(S))	IPv6 (TLS, TLS + HTTP(S))	Both (TLS, TLS + HTTP(S))
Akamai	(893, 881)	(244,241)	(227, 223)
Alibaba	(292, 175)	(38, 37)	(27, 26)
Amazon	(402, 171)	(26, 11)	(23, 7)
Apple	(502, 219)	(143, 117)	(130, 104)
Disney	(71, 2)	(3, 0)	(1, 0)
Cdnetworks	(33, 0)	(1, 0)	(0, 0)
Fastly	(8, 6)	(2, 2)	(0, 0)
Google	(5043, 4976)	(1376, 1342)	(1328, 1291)
Hulu	(39, 1)	(2, 0)	(2, 0)
Meta	(2599, 2565)	(1239, 1231)	(1195, 1185)
Microsoft	(666,174)	(47, 2)	(42, 0)
Netflix	(2917, 2731)	(977,928)	(937, 860)
Twitter	(9, 5)	(3, 2)	(2, 1)
Yahoo	(14, 0)	(2, 0)	(1, 0)

Table 5. Number of off-net ASes (TLS validated, TLS + header validated) per hypergiant.

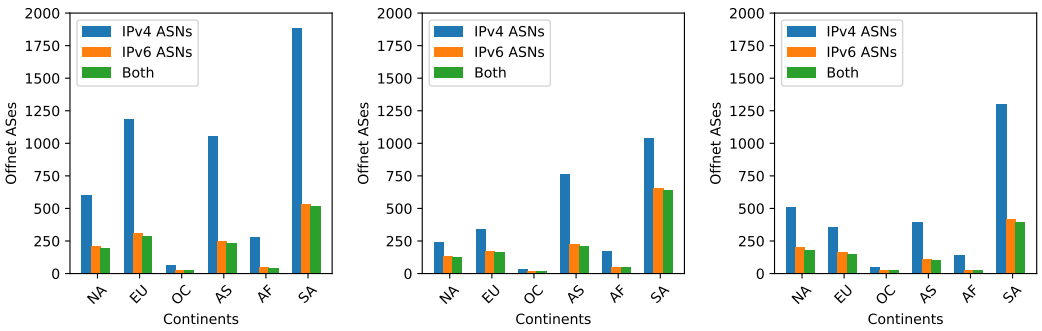
A OFF-NETS BASED ON TLS ONLY AND TLS PLUS HEADER VALIDATION

The number of IPv4 and IPv6 off-net IP we uncover across different hypergiants are detailed in Table 4. The table also presents the off-net count resulting from TLS validation (candidate off-nets) only as well as TLS plus HTTP(S) validation (final off-nets). It should be noted that the off-net IPs resulting from the later validation are always a subset of the former. We observe that counts resulting from either metric show a great degree of fluctuation across different hypergiants. Focusing on the TLS only validation, we see that the hypergiant off-nets range from under a 100 for Yahoo and Fastly to several thousands of IPs for other hypergiants such as Alibaba, Microsoft, Google, Meta, Netflix and Apple.

While these hypergiants are found to host services in ASes outside their own, we find that some of them rarely use their own infrastructure to achieve this. This can be deduced by the rapid drop in the IP numbers upon applying the HTTP(S) header validation. Hypergiants such as Microsoft,

IPv4			IPv6		
Certs (IP, ASN)	Inside HGs (IPs)	Outside HGs (IP, ASN)	Certs (IP, ASN)	Inside HGs (IPs)	Outside HGs (IP, ASN)
(27.8M, 58.6k)	2.5M	(608.7k, 14.8k)	(1.9M, 6.4k)	175.3k	(162.4k, 4.2k)

Table 6. The total number of valid certificates (sending IP addresses and ASNs), valid certificates belonging to HGs inside HG ASes and outside of HG ASes (with sending IP addresses and ASNs), over IPv4 and IPv6.



(a) Google. (b) Meta. (c) Netflix.

Fig. 9. Absolute number of off-net deploying ASes per continent for the top three hypergiants.

Disney, Apple and Amazon stand out in this with drops of 98.3%, 99.8%, 97.5%, 69.7% respectively We find this drop to result in part from the finding that some of these hypergiants use the hosting infrastructure/IPs of other hypergiants to provide their services. We discuss this in more detail in Section 4.3. In contrast, major players such as Google, Netflix, Meta, and Alibaba seem to be mostly relying on their own infrastructure for delivering their services from deployments close to their users as is reflected from similar figures from both validations. However, we confirm previous findings [4] as we also find Netflix to rely upon Amazon in part for hosting its content.

B VALID CERTIFICATES OBTAINED THROUGH MEASUREMENTS

Table 6 shows an overview of our IPv6 and IPv4 TLS scans. Over IPv6, we receive valid certificates from 1.9M IP addresses in 6.4k ASes. Nearly, 175k of these IP addresses have certificates from some hypergiant and are located outside hypergiant ASes while a lower 162.4k IP addresses with hypergiant certificates are outside in over 4k ASes. The IPv4 results show similar trends albeit with significantly larger numbers. While the number of IPv6 addresses with hypergiant certificates outside HG ASes are comparable to those within HGs, for IPv4 the IP addresses of the former latter category are nearly four times higher than the former.

C GEOGRAPHICAL DISTRIBUTION: ABSOLUTE NUMBER OF OFF-NET ASES

Figure 9 shows the absolute number of IPv6 and IPv4 off-net ASes per continent for Google, Meta, and Netflix.

D ROV AND OFF-NETS

Table 7 shows how the ROV BGP prefixes compare to BGP Off-net ROV prefixes for the off-net ASes uncovered for the top three hypergiants.

	IPv6		IPv4	
Hypergiant	ROV BGP > BGP Off. ROV	BGP Off. ROV > ROV BGP	ROV BGP > BGP Off. ROV	BGP Off. ROV > ROV BGP
Google	67.8%	32.2%	49.3%	50.7%
Meta	70.2%	29.8%	48.4%	51.6%
Netflix	64.4%	35.6%	42%	58%

Table 7. ROV analysis for the off-net ASes of the top three hypergiants

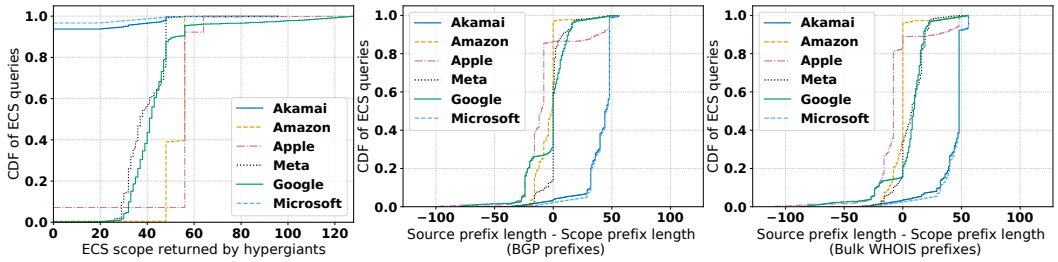


Fig. 10. (Left) ECS scope returned by the different hypergiants. (Middle and right) Difference between the source prefix length returned and the scope prefix length for BGP prefixes (Middle) and bulk WHOIS prefixes (right). A strictly negative value indicates that our prefix was not specific enough to get the most appropriate response from the hypergiant.

E ECS BEHAVIORS OF HYPERGIANTS

We look at the scope prefix length returned by the different hypergiants in their ECS responses, which specifies, according to RFC 7871, the prefix length that was expected by the hypergiants to provide the most appropriate response [18].

We first look at the distribution of the scopes returned by the hypergiants to draw a picture of the different possible behaviors. The left figure of Figure 10 shows the CDF of the different scopes returned by the different hypergiants across the DNS queries with ECS of the three datasets of prefixes (BGP, bulk WHOIS, and /56 prefixes from hitlist). First of all, Microsoft and Akamai have only 3% and 6% of the queries having a non 0 scope, showing that they do not use ECS for most of the prefixes. Then, for the remaining hypergiants, we either observe a wide range of scope in the case of Google and Meta, or only a few values for Amazon and Apple. Interestingly, we also notice that Apple has 8% and Google has 5% of their queries returning scopes that are most specific than 56, which is the maximum recommended by RFC 8781. For Google, the scope can go up to 128!

The middle and right figure of Figure 10 show the CDF of the difference between the source prefix length in the DNS query and the scope prefix length in the DNS response for the BGP prefixes and the bulk WHOIS prefixes. A strictly negative value means that the source prefix length was not specific enough to get the most appropriate response from the hypergiant, while a positive value means that the source prefix length was too specific. We exclude Akamai and Microsoft which almost always return scope prefix length of 0 from the following analysis. For BGP prefixes, the range of strictly negative values go from 13% for Meta to 86% for Apple. Similarly, for the bulk WHOIS prefixes, the range of strictly negative values go from 16% for Meta to 83% for Apple. These results show that neither BGP prefixes or the bulk WHOIS prefixes are the right granularity to use to get the appropriate ECS response. We do not show the graph for the /56 prefixes from the hitlist, because by definition, we cannot not have strictly negative values, except the few scope prefix lengths that are more specific than 56 for Apple and Google. We nonetheless mention that

the range of strictly positive values goes from less than 1% for Amazon to more than 99% for Meta, showing that always taking the /56 is also not optimal.

These results show that finding the optimal set of ECS prefixes is a hard problem and that the hypergiants adopt different strategies to respond to ECS queries.

Received December 2023; revised January 2024; accepted March 2024.