Enabling Multi-Hop ISP-Hypergiant Collaboration

Munteanu, Cristian; Gasser, Oliver; Poese, Ingmar; Smaragdakis, Georgios; Feldmann, Anja

DOI
10.1145/3606464.3606487

Publication date
2023

Document Version
Final published version

Published in
Proceedings of the Applied Networking Research Workshop

Enabling Multi-hop ISP-Hypergiant Collaboration

Cristian Munteanu
Max Planck Institute for Informatics

Oliver Gasser
Max Planck Institute for Informatics

Ingmar Poese
BENOCS GmbH

Georgios Smaragdakis
Delft University of Technology

Anja Feldmann
Max Planck Institute for Informatics

ABSTRACT

Today, there is an increasing number of peering agreements between Hypergiants and networks that benefit millions of end-user. However, the majority of Autonomous Systems do not currently enjoy the benefit of interconnecting directly with Hypergiants to optimally select the path for delivering Hypergiant traffic to their users.

In this paper, we develop and evaluate an architecture that can help this long tail of networks. With our proposed architecture, a network establishes an out-of-band communication channel with Hypergiants that can be two or more AS hops away and, optionally, with the transit provider. This channel enables the exchange of network information to better assign requests of end-users to appropriate Hypergiant servers. Our analysis using operational data shows that our architecture can optimize, on average, 15% of Hypergiants' traffic and 11% of the overall traffic of networks that do not interconnect with Hypergiants. The gains are even higher during peak hours when available capacity can be scarce, up to 46% for some Hypergiants.

CCS CONCEPTS

• **Networks** → **Network architectures**.

KEYWORDS

Internet Architecture, Content Delivery, Traffic Optimization.

ACM Reference Format:
Cristian Munteanu, Oliver Gasser, Ingmar Poese, Georgios Smaragdakis, and Anja Feldmann. 2023. Enabling Multi-hop ISP-Hypergiant Collaboration. In *Applied Networking Research Workshop (ANRW '23), July 24, 2023, San Francisco, CA, USA.* ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3606464.3606487

1 INTRODUCTION

More than half of the traffic delivered to Internet users originates from a small number of content providers, referred to as Hypergiants (HGs) [16, 18, 24]. Hypergiants (e.g., Google, Netflix, Facebook) are content or cloud providers, and content delivery networks that have a heavy out-bound traffic profile (i.e., they send much more traffic then they receive) [4, 10, 16]. They operate thousands of servers [10], and generally do not offer Internet access to end users.

Over the years, Hypergiants have evolved to some of the largest global networks. To cope with the unprecedented demand and to improve end-user experience and engagement, they have established interconnections with thousands of networks. For example, Google has established interconnections with more than 15 thousand networks [3] and Cloudflare claims that it has established interconnections with more than 10 thousand networks [8]. Although this "flattening" of the Internet topology has benefited thousands of networks (typically large eyeball networks) and millions of Internet users, there is a long tail of more than 40 thousand networks with less privileged Internet users. This is of concern, as there are no economically sustainable models to interconnect all the Hypergiants with all these (typically small eyeball or enterprise) networks. Indeed, most of the Hypergiants require a minimum level of traffic to establish a direct interconnection. For example, Google's traffic requirement for private peering is 1 Gbps [13]. Public peering at IXPs offers scalability and the opportunity for smaller networks to establish peerings with Hypergiants, typically without minimum exchange traffic level [13, 19]. However, many small networks like regional eyeball network and enterprises may not have easy access to an IXP nor can afford personnel and hardware cost to be present and operate in a colocation center or an IXP. Such limitations will lead to networks being left behind as they can not benefit from the increasing interconnection between Hypergiants, which in turn can potentially contribute to a digital divide [6, 11].

Our motivation for this work is that even though Hypergiants have direct interconnections or co-hosting arrangements with many ISPs they do not have these with all ASes. Indeed, from our work with two major transit ISPs we know of at least 20 customer ASes that have multiple interconnections to these ISPs in different locations (datacenters). While we acknowledge that most of today's Internet traffic does not use multiple AS hops [7] the traffic volume is still substantial and rather important for each of the individual smaller ASes. As such it is important to find ways that they can influence the selected servers or the chosen paths. Our contributions can be summarized as follows:

- We present the architecture and evaluation of a system that enables collaboration between Hypergiants and remote networks without establishing direct peering. We show different versions of our system that involve different pairs of parties: Hypergiants, transit and customer networks.
- Our analysis with operational traffic shows that up to 15% of the Hypergiant traffic that is delivered to end users can be optimized with no additional arrangements by transit providers, customer networks, or Hypergiants.

ANRW '23, July 24, 2023, San Francisco, CA, USA

ACM ISBN 979-8-4007-0274-7/23/07.
https://doi.org/10.1145/3606464.3606487

- We show that the optimized traffic can be up to 28% during peak time when link capacity may be a scarce resource.
- Our analysis also shows that the optimization gains may differ across Hypergiants. For some, the optimization of their traffic can be up to 46% of traffic volume delivered to customer ASes.

2 BACKGROUND

CDN server selection: Every time a user accesses content from a Hypergiant the Hypergiant has to select a server to serve the content, unless it is relying on anycast (where the selection is determined by routing). Typically, this server selection is done when the client issues the DNS request for resolving the hostname to an IP. Thus, the Hypergiant can use information about the client, i.e., via the source IP address of the DNS request or the client prefix contained in the EDNS Client Subnet (ECS) extension [1, 5, 14, 17, 21]. The Hypergiant combines this with its local information, e.g., server load, connection cost, routing information, to make an informed decision. However, neither of these information sources is accurate, e.g., the resolver IP address can be misleading if the host is using a public DNS resolver [12] and the routing information via BGP announcements may be too coarse grained [20, 26].

ISP-HG collaboration: Lack of information about a user's network location can lead to a non-optimal server selection and non-optimal path choices within the ISP [12, 15]. This challenge can be addressed by exchanging information between the involved parties. So far, this has been proposed for directly interconnected parties only. Solutions that share maps between two parties are *ALTO* [2] and *P4P* [25]. A different approach is *FlowDirector* [18] where a dedicated server within the ISP collects network information to maintain the latest state of the network activity. The difference between FlowDirector and ALTO or P4P is that the latter provide an information exchange while FlowDirector provides up-to-date network view service that the Hypergiant contacts to map users to appropriate servers.

3 ARCHITECTURE

Parties and their goals: Our scenario involves three types of parties each with their own goals: (I) The customer AS wants to improve where, i.e., on which border router, it receives traffic from a Hypergiant for its end-users to improve their experience. (II) The transit AS may want to reduce its cost for transiting the Hypergiant traffic through its network before delivering it to is customer ASes. (III) The Hypergiant may want to select a server close to the customer AS to improve its user experience and engagement.

Example setting: Figure 1 highlights the possible choices of each of the parties and how collaboration can lead to a better solution. Each of the ASes has two different locations, 1 and 2. Routers A1, B1, C1, and D1 are in location 1 and routers A2, B2, C2, and D2 in location 2.

Once the client within the customer AS requests a resource from the Hypergiant, the Hypergiant uses the DNS request to map it to one of the two server deployments, e.g., servers $S1$ in Figure 1 (a). Using the underlying routing system the traffic is forwarded via the green route. Yet, this may not be optimal for either the end-user, the customer AS, nor the transit AS. Without changing the server selection the transfer of data from location 1 to location 2 cannot

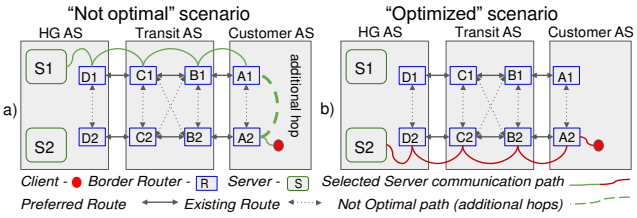

Figure 1: Server selection for 2-hop scenario showing the problem of a non-optimally selected server. Routers A1, B1, C1, D1 are present at Location 1. Routers A2, B2, C2, D2 are present at Location 2. a) server is selected leading to a non-optimal path. b) optimal solution.

be avoided. Yet, if the Hypergiant chooses $S2$, all involved parties are likely to benefit, see the red route in Figure 1 (b).

One may argue that using correct geolocation information may be sufficient to make the best choice. However, this is not always the case as the path between B2 and C2 or the peering link between A2 and B2 may be congested. In this case Figure 1 (a) may be the "optimal" solution. Moreover, geolocation is often imprecise [22] and short physical distances do not necessarily result in short topological distances, in particular, if there is another AS involved in between. Thus, we argue that the parties should explicitly exchange information between them. If there are additional transit ASes involved, the setting becomes more complex. As such we focus on the outlined three party collaboration.

Collaboration scenarios: Our settings involve three entities that should communicate with each other. Hereby, we only assume that the transit AS and the customer AS already have an agreement which is likely since the transit AS is the provider of the customer AS. The transit AS may have an agreement with the Hypergiant to collaborate [18], but this is no requirement. As such we consider three possible collaboration scenarios: (I) Multi-hop: Here, the communication occurs directly between the customer AS and the Hypergiant. As such the transit ASes are not involved. (II) One+-hop: In this case, the communication is steered via the transit ASes. Transit ASes may want to offer this as a service to their customer ASes to increase engagement and attract new customers. Transit providers may also want to minimize any side effect of out-of-band communication between Hypergiants-customers, e.g., congestion in the (transit) network. In this case the transit ASes have to be the intermediate between Hypergiants and customer ASes. (III) Full: Here, all parties exchange information. This is similar to One+-hop, but there is also a direct out-of-band communication between the Hypergiants and customer ASes.

Note, the goal of the proposal is to exchange information that can be used to improve traffic steering. However, none of the parties has to act on any of the information nor does a party have to provide all of the information. Still, the benefit is the largest if all parties actively participate.

The customer AS sends its prefixes and preferences. The expression of preferences can take place in two ways. The first way is to use a similarity expression per destination prefix, e.g., traffic for prefix P_1 should be received on the same interface as traffic for prefix P_2 or P_3, in this order of preference. The second way is to refer to specific network interfaces on which an AS wants to receive

traffic for a destination prefix, e.g., traffic for prefix P_1 should be received via next hop IP_1 or IP_2 or IP_3, in this order of preference. The advantage of the latter is that it is precise and easy to specify. However, it only applies for communication between two directly connected ASes. The advantage of the former is its generic nature. However, it needs at least one reference prefix where the traffic flow works as desired. We propose to use the first way for the Multi-hop scenario and the second way for the One+-hop. The Full scenario can use either one or even a mixture of the two ways.

4 SYSTEM IMPLEMENTATION

Given the three involved parties we split the tasks into three modules—one for each of the different types of ASes.

Customer AS module: This module selects relevant prefixes along with preferences and sends the information to the next transit AS or the Hypergiant. The module is used by the customer AS to create a dictionary of key:value pairs for each Hypergiant. The key is the prefix and the value is the list of preferred next hops or the list of "similar" prefixes. This module has to address the following two challenges: (i) How to select the client prefixes and (ii) how to determine the best next hops (or similar prefixes) for each prefix. The involved steps are as follows: (1) If the DNS resolvers of the customer AS and the Hypergiant support the EDNS Client Subnet (ECS) extension [23], the module identifies prefixes at appropriate granularities. If there is no support for ECS the module focuses on the DNS resolvers. Recall, the Hypergiant has to use the DNS resolvers IP as the basis for its resolver selection process. Thus, we group client prefixes by which DNS resolver they use by default. Note that if one DNS resolver fails and the clients are remapped to a different one this mapping has to be updated quickly. (2) Next, we determine for each of the prefixes, the preferred next hops for the ingress traffic using the current network topology (collected, e.g., from IGP data). This results in an initial list of tuples of either: ("prefix":list of "similar prefixes"), ("prefix":list of "preferred next hops"), or a combination between the two. (3) Next, we decide on appropriate prefix length. This can either de-aggregate some of the prefixes to give more freedom or aggregate them further to avoid redundancy.

Transit AS module: This module receives the prefix preferences from its customers (which can also be another transit AS). Based on its topology and traffic flow, it determines its own preference for each aggregated or de-aggregated prefix and sends this information onward. Note that this module may not be required in a Multi-hop scenario. The involved steps are as follows: (1) The transit AS receives a list of key:value tuples per Hypergiant from each participating customer AS. (2) It then can aggregate this information and identify for each Hypergiant and each prefix the best next hop within its network using its current network topology. This results again in a list of tuples of either: ("prefix":list of "similar prefixes"), ("prefix":list of "preferred next hops") or a combination between the two. (3) Before sending the dictionary the transit AS has again the choice of either aggregating or de-aggregating prefixes or even deleting or adding prefixes. Note, the transit AS's interests may not always match with the downstream AS's interest, e.g., the customer AS. There can be conflicts, whereby, the transit AS has the choice of either optimizing its traffic flow or choosing the most desired next hop of the downstream.

Hypergiant AS module: This module receives the preferences from either the customer AS or the transit AS and can use them to refine its server selection process. The involved steps are as follows: (1) The Hypergiant receives the dictionary and can use the information in its server selection process. (2) To further optimize the server selection process, the Hypergiant and the customer AS can agree to, e.g., increase the support for ECS, deploy additional DNS servers (even if they are just virtual ones), or agree on a different prefix aggregation level.

5 DATASETS

We obtain a week of opearational data (November 1-7, 2021) by establishing collaborations with network providers.

ISP topology. The data was gathered from internal routing information at a large European transit network.

ISP traffic data. We obtain the egress traffic captured from all border routers from a large European network as well as multiple of its customers. Data was collected with IPFIX with a consistent sampling rate identical at all the border routers. From the ingress traffic it was possible to infer the traffic flow from Hypergiants to ISP's customer networks.

BGP data. We obtain matching BGP data from the peering routers of the ISP.

Ethical considerations. Our study is based on traffic and topology data that the ISPs regularly capture for operational purposes and are in compliance with legal requirements in the respective countries of operation. All traffic traces are aggregated at flow-level and, thus, do not contain any payload. Additionally, the data is processed and analyzed in-situ at the premise of the ISPs. We anonymize all networks and normalize traffic volume in the study to comply with the requirements set forth by the collaborating companies.

6 EVALUATION: POTENTIAL BENEFITS

Our collaborations allowed us to get access to data from a Tier-2 AS (AS-C) which serves roughly 11 million customers and its major transit provider, Large European Transit AS (AS-T). The Large European Transit AS is responsible for roughly 60% of the ingress traffic. The Tier-2 AS also has direct interconnections with two Hypergiants, namely, HG12 and HG13. HG12 is responsible for roughly 37% of the ingress traffic while HG13 is responsible for roughly 1.4%. The remaining 1.6% of traffic are ingressing from other non-Hypergiant peers. AS-C and AS-T are interconnecting in two locations. AS-C operates a single router at each location but interconnects to AS-T via multiple interconnection links, i.e., it has multiple next hop candidate links at each location.

Our initial goal is to understand the potential benefit of the one+-hop ISP-Hypergiant collaboration. For this we only focus on the transit links between AS-T and AS-C. We note, that none of the two Hypergiants with direct interconnections receive any traffic via the peering link. As such we do not consider the special case where the customer AS also has direct interconnection links with the Hypergiants.

The first step is to identify the prefix ranges of the Hypergiants that are two hops away from our customer AS, AS-C. We select

Hypergiant	Traffic %	Not-optimized %	Not-optimized % per own traffic share
HG1	31.93%	0.59%	1.86%
HG2	16.17%	2.97%	18.38%
HG3	8.15%	1.78%	21.90%
HG4	6.96%	3.21%	46.15%
HG5 *	4.46%	1.70%	38.10%
HG6	3.09%	1.07%	34.62%
HG7	2.62%	0.06%	2.27%
HG8	2.26%	0.24%	10.53%
HG9	2.26%	0.78%	34.21%
HG10 *	2.08%	0.75%	36.00%
HG11 *	2.08%	0.76%	37.00%
Others	17.95%	—	—
Total	**100%**	**13.91%**	

Table 1: Summary of traffic shares per Hypergiant. "*" indicates that server selection is not sufficient to optimize the traffic flow. Here, the Large European Transit AS has to change its routing to help the Tier-2 AS.

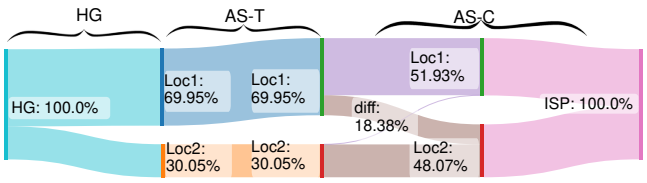

Figure 2: Sankey diagram of traffic flow between HG2 via the Large European Transit AS to the Tier-2 AS.

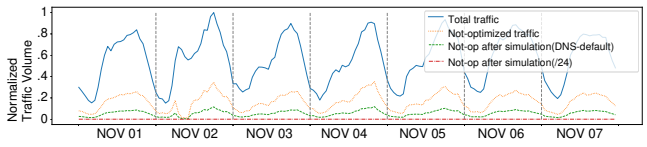

Figure 3: HG2 traffic flows. The large fraction of not-optimized traffic highlights the potential benefits of collaboration and the minimal fraction of not-optimized traffic highlights that collaboration can achieve the potential.

the top 15 Hypergiants [4, 10], identify their AS number, and check which of these are peering with AS-T. All but three HG ASes peered with AS-T. Note, AS-T does not host any Hypergiant infrastructure within their network. Next, we use the Hypergiant AS numbers and the BGP data to identify the prefixes of the Hypergiants and their traffic contributions. In addition, we use the topology data to determine the traffic share per path from the Hypergiant to the AS-C within AS-T. Using this data we can assess the potential benefits of the multi-hop collaboration.

Table 1 summarizes our observations. Our analysis shows that the top 15 Hypergiants contribute roughly 82% of the total traffic. Of this traffic 13.91% is not-optimized. Here not-optimized refers to two different scenarios. The first scenario corresponds to the example shown in Figure 1 (a). The Hypergiant has interconnections in both locations where AS-T and AS-C interconnect and the Hypergiant is sending the data via the "wrong" location. In the second scenario, the Hypergiant is peering with the Large European Transit AS in one or multiple locations but not both of the locations. In these cases, all the traffic is sent to AS-C in one of the locations. Cross-checking the routing shows that the chosen location is the one that is optimal for the Large European Transit AS. However, it may not be optimal for the Tier-2 AS. Here, AS-T has the potential to help AS-C in the one+-hop or the full cooperation cases.

Overall, the latter accounts for traffic of three Hypergiants and 3.21% of not-optimized traffic while the former applies to traffic of eight Hypergiants and results in 10.7% of not-optimized traffic relative to the total traffic volume. We note that the not-optimized fraction of traffic differs substantially per Hypergiant and ranges from 1.86% to 46.15% of its traffic share. Note, the latter share does not include any additional optimization that the Large European Transit AS may do to help the Tier-2 AS. In the remainder of our analysis we do not focus on the case where the Large European Transit AS has to be willing to help the Tier-2 AS. As such we focus on the 73.43% of total traffic where the main benefit of the collaboration comes from the Hypergiant server selection.

7 EVALUATION: CASE STUDY HG2

Given the potential improvements we next focus on one of the Hypergiants, namely, HG2. Our motivation for choosing HG2 is four-fold: (a) the Large European Transit AS has already an established one-hop cooperation with the Hypergiant for improving its traffic flow; (b) HG2 supports ECS which simplifies the prefix aggregation/de-aggregation study; (c) HG2 is willing to experiment on the one+-hop cooperation schema; (d) HG2 has a larger unoptimized traffic ratio than HG1.

Figure 2 shows the potential improvements for this Hypergiant as a Sankey diagram. It shows what fraction of traffic is flowing from the Hypergiant (on the left) to the customer AS, AS-C (on the right). We see that the Hypergiant traffic is split between the two input location of AS-T in a ratio of around 70% to 30%. Within AS-T the traffic is forwarded without issues. However, within AS-C the traffic has to be redistributed. Only 51.93% of the traffic entering AS-C have their best path close to this location. This implies that 18.2% of the traffic has to be rerouted within the Tier-2 AS for this location. Moreover, of the 30% entering AS-C in the other location 0.18% of the traffic has to be rerouted as well. Overall, 18.38% of the HG2 traffic are using servers that are non-optimally selected.

Next, we check how the traffic volume and the non-optimally selected traffic volume for this Hypergiant is changing across time. The blue solid line in Figure 3 shows this for the whole one-week period. The traffic volume is normalized by the maximum observed traffic volume. The plot shows the typical time of day effects as in other residential ISPs [9]. The traffic volume increases throughout the work days with plateaus during lunchtime and peaks during the evening hours. During weekend days the traffic volume increases earlier. Overall, we see that the fraction of not-optimized traffic (orange dotted line) matches this trend but is even higher during busy hours. This indicates that the optimization potential is higher than the average numbers may indicate. For example, it reaches more than 24% during busy hours, when capacity may be a scarce resource.

On November 2, we see a substantial drop in the not-optimized traffic that covers a two-hour period. This is the result of a maintenance activity by HG2. During this period servers close to one location were not available and no clients were assigned to them. Thus, most client requests were served by servers at different locations. These server locations were "accidentally" better suited for the clients as they resulted in an optimized route. After the maintenance period, the not-optimized traffic returned to its old fraction. This event underlines that small changes in server selection can result in substantial improvements for the customer AS. To calculate the improvement that can be potentially achieved via ISP-Hypergiant collaboration we assume that the Hypergiant will benefit and thus use the additional information received for server selection. In our simulation we follow the steps outlined in Section 4.

While doing this, we identify 8 prefixes announced by the Tier-2 AS which corresponds to 273 /24 prefixes. We confirm that the best path is via the Large European Transit AS and that the AS-PATH length is two. To map prefixes to DNS resolvers we rely on the traffic captures which also contain the DNS requests them self. To get a one-to-one mapping of prefixes to DNS resolvers we de-aggregate some of the prefixes. This results in roughly 70 prefixes. We then use the topology information to determine the next hop that is currently used as well as the "optimum" one (in terms of latency) for each prefix. This corresponds to identifying the optimum path. We note, that the Tier-2 AS uses—for some of the prefixes—smaller prefixes length in their internal routing. As such there may be multiple next hop candidates. In such a case, we use the one that is chosen by the majority of the traffic. This results in some "suboptimal" assignment of traffic. To check how much, we later study the impact of further de-aggregating prefixes. All this information is then aggregated in a preference dictionary and send to the transit AS. We note, that neither the assignment of prefixes to DNS servers nor the internal routing changed during the week that we studied.

For the transit AS we use its topology information as well as its routing policies to check which border router would be the optimal one (in terms of latency) for each of the prefix. We also check that each of the path has sufficient spare capacity that they can be chosen. Note, we optimize for latency as it also reduces the number of long distance links that are chosen. This is likely to reduce the cost both for the transit AS as well as the customer AS. This information is then sent to the Hypergiant. The impact on the traffic flow when the Hypergiant follows the recommendations are shown in Figure 3, see the green dashed line for not-optimized traffic flow with one+-hop/multi-hop collaboration. Overall, we see that the positive effects of the collaboration. All but 1.37% of the traffic is not-optimized, i.e., not routed "optimally". The reason for the remaining small percentage of not-optimal traffic is our choice of prefix de-aggregation. The "default" de-aggregation that we choose in this case is the mapping of prefixes to DNS servers determined by the Tier-2 AS. However, as the Tier-2 AS is using a different prefix aggregation for its internal routing this is not always the best choice.

To assess the impact of choosing a specific maximum prefix length on the fraction of not-optimized traffic we ran a series of simulations. We varied the maximum chosen prefix length from

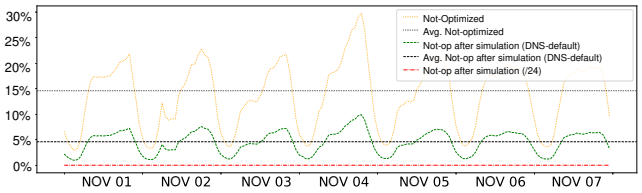

Figure 4: All Hypergiants: ratio of not-optimized traffic flows vs. total traffic flow. This highlights that the potential benefits are even larger during busy hours.

Hypergiant	Original Not-opt	Not-opt after Simulation	
	BGP ann. (#prf.)	'/24' (#prf.)	DNS-default (#prf.)
HG1	1.86% (8)	0% (371)	1.86% (69)
HG2	18.38% (8)	0% (273)	1.37% (70)
HG3	21.90% (8)	0% (268)	11.44% (62)
HG4	42.80% (8)	0% (182)	8.93% (40)
HG6	34.62% (8)	0% (145)	15.44% (28)
HG7	2.27% (8)	0% (144)	2.27% (25)
HG8	10.53% (8)	0% (138)	7.62% (24)
HG9	34.21% (8)	0% (132)	6.21% (24)

Table 2: Per Hypergiant percentage of not-optimized traffic: with original prefix announcements; with high disaggregation, i.e., "/24" prefixes; with prefix disaggregation.

"/16" (which results in 12 prefixes) up to "/24" (resulting in more than 200 prefixes). Using "/16" does not improve the server selection at all since it does not give the Hypergiant enough information. Indeed, it may result in a worse server selection. Once we reach "/20" most of the potential benefit has been leveraged. Note, "/20" also corresponds to the inferred mapping of prefixes to DNS servers; we refer to this as "DNS-default". Still, by further de-aggregating, i.e., up to "/24", it is possible to reduce the fraction of not-optimized traffic to 0%.

This simulation assumes no changes to BGP or in the internal routing of the Tier-2 AS nor the Large European Transit AS. We also checked that the link as well as the router capacities are sufficient. Thus, purely by changing the server selection and consequently the source of the traffic, we can reduce traffic on long distance links by more than 18%. These achievements apply to the one+-hop case as discussed above but also to the multi-hop case.

8 HYPERGIANT OPTIMIZATIONS

Next, we analyze if the potential benefits, see Section 6, are achievable for the other Hypergiants in a similar fashion as it is the case for HG2, see Section 7. Recall, that each Hypergiant has a different fraction of not-optimized traffic, ranging from 46.15% for HG4 to 1.86% for HG1. Such large differences may be the result of different operational practices or different priorities within the Hypergiants, e.g., [20, 26].

We run the same simulations for the other seven Hypergiants where optimization is possible. We exclude three Hypergiants where changing the server selection alone does not result in any improvements in the traffic flow for the Tier-2 AS. The results are shown in Table 2. We note that using the original BGP prefixes for the ISP-Hypergiant collaboration does not yield substantial benefits. It only improves the fraction of non-optimized traffic for HG4 by a

small fraction. None of the other Hypergiants see improvements. However, using a large de-aggregation to "/24s" eliminates all non-optimized traffic for all Hypergiants. In this case, the number of advertised prefixes also increases from 8 to tens, and for large Hypergiants up to 370. Using the DNS-default prefix de-aggregation (based on which DNS local resolver is used) shows substantial improvements and may often be sufficient. Still, for HG1 and HG7 using the DNS-default one does not show any noticeable improvements. Here, further de-aggregation is needed. However, the degree of de-aggregation differs by Hypergiant. For some, 132 prefixes up to "/24s" are sufficient for others more than 350 are needed. Overall, the share of unoptimized overall traffic is reduced from 14.6% to 4.6%.

Figure 4 shows the ratio of non-optimized traffic vs. total traffic (orange dotted line) across time as well as the average ratio (black dotted line). Looking at the fraction of not-optimized traffic using DNS-default prefix de-aggregation (green dashed line) we see that the total benefit is smaller than for the case of HG2. This is in line with the averages shown in Table 1. Still, during busy hours the benefits increase, see the green and black dashed lines in Figure 4. Yet, the differences while visible are not dominating. Using full de-aggregation to "/24s" allows us to take full advantage of the Hypergiant-ISP collaboration and we can eliminate not-optimized traffic. This results in the flat red dashed/dotted line at zero shown in Figure 4. Notice that the maximum benefit is during the peak hour, when the link capacity may be a scarce resource. The optimization gain compared to the original case during peak time is up to 28%.

9 CONCLUSION

We present the architecture and evaluation of a system that enables collaboration between Hypergiants (HGs) and remote networks without establishing direct peering. We present different versions of our system that involve pairs of parties: HGs, transit and customer networks. Data-drven analysis shows that up to 15% of the HG traffic delivered to a two-hop remote customer ASes can benefit. The gains are even higher (28%) during peak time. For some HGs, up to a third of their traffic delivered to users of remote customer networks can be optimized with no changes on the peering relationships of HGs, transit, and customer ASes. We are currently in contact with customer ASes,

HGs and transit providers that are interested in testing our solution. As part of our future research agenda, we will report on our experience in operating the multi-hop and one+-hop solution for multiple Hypergiants, transit, and customer networks around the globe.

ACKNOWLEDGMENT

This work was supported in part by the European Research Council (ERC) Starting Grant ResolutioNet (ERC-StG-679158).

REFERENCES

[1] R. Al-Dalky, M. Rabinovich, and K. Schomp. 2019. A Look at the ECS Behavior of DNS Resolvers. In *ACM IMC*.
[2] R. Alimi, R. Penno, and Y. Yang. 2011. ALTO Protocol. IETF RFC 7285. (2011).
[3] T. Arnold, J. He, W. Jiang, M. Calder, I. Cunha, V. Giotsas, and E. Katz-Bassett. 2020. Cloud Provider Connectivity in the Flat Internet. In *ACM IMC*.
[4] T. Böttger, F. Cuadrado, and S. Uhlig. 2018. Looking for Hypergiants in PeeringDB. *ACM CCR* 48, 3 (2018).
[5] M. Calder, X. Fan, and L. Zhu. 2019. A Cloud Provider's View of EDNS Client-Subnet Adoption. In *Network Traffic Measurement and Analysis Conference (TMA)*.
[6] I. Castro, J. C. Cardona, S. Gorinsky, and P. Francois. 2014. Remote Peering: More Peering without Internet Flattening. (2014).
[7] N. Chatzis, G. Smaragdakis, J. Boettger, T. Krenc, and A. Feldmann. 2013. On the benefits of using a large IXP as an Internet vantage point. In *ACM IMC*.
[8] Cloudflare. 2022. Project Myriagon: Cloudflare Passes 10,000 Connected Networks. https://blog.cloudflare.com/10000-networks-and-beyond/. (2022).
[9] A. Feldmann, O. Gasser, F. Lichtblau, E. Pujol, I. Poese, C. Dietzel, D. Wagner, M. Wichtlhuber, J. Tapiador, N. Vallina-Rodriguez, O. Hohlfeld, and G. Smaragdakis. 2021. A Year in Lockdown: How the Waves of COVID-19 Impact Internet Traffic. *Communications of the ACM* 64, 7 (July 2021).
[10] P. Gigis, M. Calder, L. Manassakis, G. Nomikos, V. Kotronis, X. Dimitropoulos, E. Katz-Bassett, and G. Smaragdakis. 2021. Seven Years in the Life of Hypergiants' Off-Nets. In *Proc. ACM SIGCOMM*.
[11] V. Giotsas, G. Nomikos, V. Kotronis, P. Sermpezis, P. Gigis, L. Manassakis, C. Dietzel, S. Konstantaras, and X. Dimitropoulos. 2021. O Peer, Where Art Thou? Uncovering Remote Peering Interconnections at IXPs. *IEEE/ACM Transactions on Networking* 29, 1 (2021).
[12] U. Goel, M. P. Wittie, and M. Steiner. 2015. Faster Web through Client-Assisted CDN Server Selection. In *24th International Conference on Computer Communication and Networks (ICCCN)*.
[13] Google. 2022. Google Peering Technical Requirements. (2022). https://peering.google.com/#/options/peering
[14] A. Kountouras, P. Kintis, A. Avgetidis, T. Papastergiou, C. Lever, M. Polychronakis, and M. Antonakakis. 2021. Understanding the Growth and Security Considerations of ECS. In *NDSS*.
[15] M. Kwon, Z. Dou, W. Heinzelman, T. Soyata, H. Ba, and J. Shi. 2014. Use of Network Latency Profiling and Redundancy for Cloud Server Selection. In *IEEE 7th International Conference on Cloud Computing*.
[16] C. Labovitz, S. Lekel-Johnson, D. McPherson, J. Oberheide, and F. Jahanian. 2010. Internet Inter-Domain Traffic. In *Proc. ACM SIGCOMM*.
[17] G. Moura C. M., S. Castro, W. Hardaker, M. Wullink, and C. Hesselman. 2020. Clouding up the Internet: how centralized is DNS traffic becoming?. In *ACM IMC*.
[18] E. Pujol, I. Poese, J. Zerwas, G. Smaragdakis, and A. Feldmann. 2019. Steering Hyper-Giants' Traffic at Scale. In *Proc. ACM CoNEXT*.
[19] P. Richter, G. Smaragdakis, A. Feldmann, N. Chatzis, J. Boettger, and W. Willinger. 2014. Peering at Peerings: On the Role of IXP Route Servers. In *ACM IMC*.
[20] B. Schlinker, H. Kim, T. Cui, E. Katz-Bassett, H. V. Madhyastha, I. Cunha, J. Quinn, S. Hasan, P. Lapukhov, and H. Zeng. 2017. Engineering Egress with Edge Fabric: Steering Oceans of Content to the World. In *Proc. ACM SIGCOMM 2017*. 418–431.
[21] K. Schomp, T. Callahan, M. Rabinovich, and M. Allman. 2013. On Measuring the Client-Side DNS Infrastructure. In *ACM IMC*.
[22] Y. Shavitt and N. Zilberman. 2011. A Geolocation Databases Study. (2011).
[23] F. Streibelt, J. Boettger, N. Chatzis, G. Smaragdakis, and A. Feldmann. 2013. Exploring EDNS-Client-Subnet Adopters in your Free Time. In *ACM IMC 2013*. 305–312.
[24] M. Trevisan, D. Giordano, I. Drago, M. M. Munafo, and M. Mellia. 2018. Five Years at the Edge: Watching Internet from the ISP Network. In *Proc. ACM CoNEXT*.
[25] H. Xie, Y. R. Yang, A. Krishnamurthy, Y. G. Liu, and A. Silberschatz. 2008. P4P: Provider Portal for Applications. In *Proc. ACM SIGCOMM*.
[26] K-K. Yap, M. Motiwala, J. Rahe, S. Padgett, M. Holliman, G. Baldus, M. Hines, T. Kim, A. Narayanan, A. Jain, V. Lin, C. Rice, B. Rogan, A. Singh, B. Tanaka, M. Verma, P. Sood, M. Tariq, M. Tierney, D. Trumic, V. Valancius, C. Ying, M. Kallahalla, B. Koley, and A. Vahdat. 2017. Taking the Edge off with Espresso: Scale, Reliability and Programmability for Global Internet Peering. In *Proc. ACM SIGCOMM 2017*. 432–445.