



STEERING HYPER-GIANTS' TRAFFIC AT SCALE

Eric Pujol¹ I. Poese¹ J. Zerwas² G. Smaragdakis³ A. Feldmann⁴



BENOCS¹



TU München²



TU Berlin³



Max Planck Inst. Informatics⁴



What are *hyper-giants*?¹²

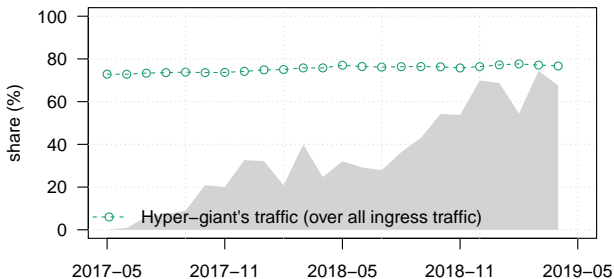
- Large networks providing services
- Global infrastructure
- Generate enormous amounts of traffic

Some of them...



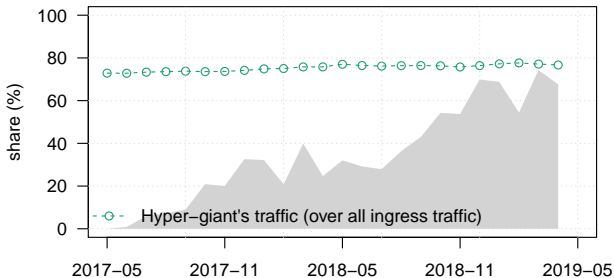
¹Labovitz et. al. "Internet Inter-Domain Traffic" in SIGCOMM'10

²Böttger et. al. "Looking for hypergiants in peeringDB." ACM CCR 48.3



A large ISP's perspective:

- > 50 million customers
- > 50 PB (daily)
- > 10 PoPs



A large ISP's perspective:

- > 50 million customers
- > 50 PB (daily)
- > 10 PoPs

Overall ingress traffic:

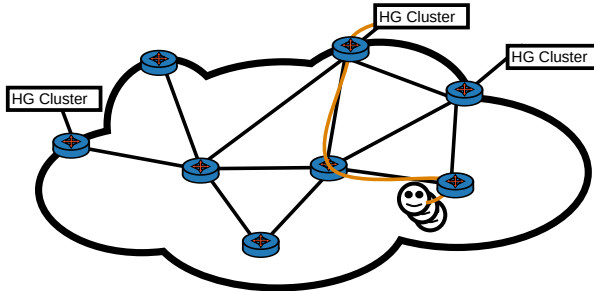
- ~ 30% growth per annum

Top 10 hyper-giants:

- ~ 75% share



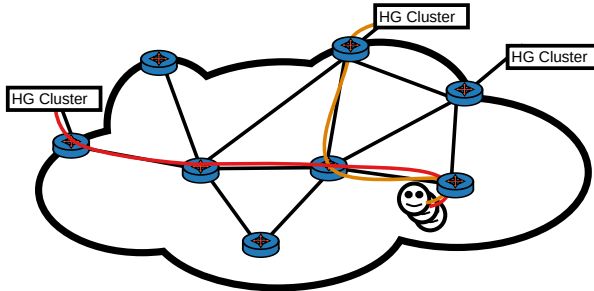
Toy example



Baseline: 2 bytes in the backbone per ingress byte



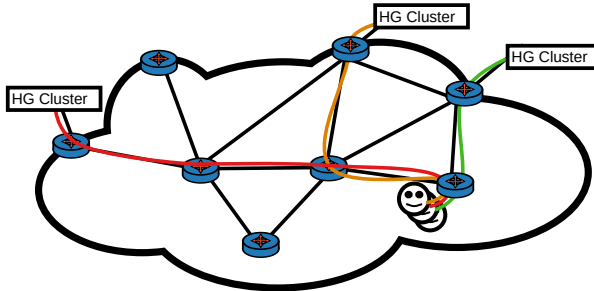
Toy example



“Bad” mapping= higher costs and incr. latency



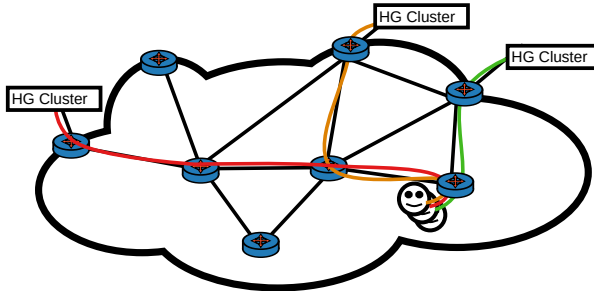
Toy example



“Better” mapping= 50% reduction



Toy example



Wait a second... This seems familiar...



Improving Content Delivery with PaDIS

Ingmar Poesche T-Labs/TU Berlin ingmar@net.t-labs.tu-berlin.de	Benjamin Frank T-Labs/TU Berlin bfrank@net.t-labs.tu-berlin.de	Bernhard Ager T-Labs/TU Berlin bernhard@net.t-labs.tu-berlin.de
Georgios Smaragdakis T-Labs/TU Berlin georgios@net.t-labs.tu-berlin.de	Steve Uhlig T-Labs/TU Berlin steve@net.t-labs.tu-berlin.de	Anja Feldmann T-Labs/TU Berlin anja@net.t-labs.tu-berlin.de

Abstract

Today, a large fraction of Internet traffic is originated by Content Delivery Networks (CDNs). To cope with the increasing demand for content CDNs, deploy massively distributed infrastructures. Moreover, to minimize their cost, content delivery networks perform their own traffic optimization by assigning end-users to their servers. Such an assignment is at large unaware of the network conditions and based on inaccurate information on the location of the end-user. Thus, users are not always assigned to the CDN servers that provide optimal end-user performance. To improve user assignment especially from a performance perspective we propose and deploy a Provider-aided Distance Information System (PaDIS). PaDIS is a novel system that allows ISPs to

more than 50% of the traffic [8, 10, 14, 4]. Among the major causes for the current prevalence of HTTP traffic, we find the increase of streaming content, e.g., offered by `youtube.com`, as well as the popularity of the content offered by One-Click Hosters (OCHs) [2] such as `rapidshare.com`. This popular content is hosted by the new "Hyper Giants" [8] which include large content providers (CPs), such as Google and Yahoo!, as well as Content Distribution Networks (CDNs), such as Akamai and Limelight [6]. To keep the terminology simple, we refer to different types of players in the content delivery landscape, e.g., CPs, CDNs and OCHs, simply as CDNs.

To achieve high levels of performance and scalability, CDNs rely on distributed infrastructures. Some of them even have



Improving Content Delivery with PaDIS

Ingmar Poesse
T-Labs/TU Berlin
ingmar@net.t-labs.tu-berlin.de

Benjamin Frank
T-Labs/TU Berlin
bfrank@net.t-labs.tu-berlin.de

Bernhard Ager
T-Labs/TU Berlin
bernhard@net.t-labs.tu-berlin.de

Georgios Smaragdakis
T-Labs/TU Berlin
georgios@net.t-labs.tu-berlin.de

Steve Uhlig
T-Labs/TU Berlin
steve@net.t-labs.tu-berlin.de

Anja Feldmann
T-Labs/TU Berlin
anja@net.t-labs.tu-berlin.de

Steering Hyper-Giants' Traffic at Scale

Enric Pujol
BENOCS
epujol@benocs.com

Ingmar Poesse
BENOCS
ipoese@benocs.com

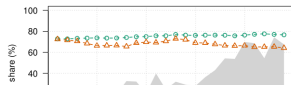
Johannes Zerwas
TU München
johannes.zerwas@tum.de

Georgios Smaragdakis
TU Berlin
georgios@inet.tu-berlin.de

Anja Feldmann
Max Planck Institute for Informatics
anja@mpi-inf.mpg.de

ABSTRACT

Large content providers, known as *hyper-giants*, are responsible for sending the majority of the content traffic to consumers. These *hyper-giants* operate highly distributed infrastructures to cope with the ever-increasing demand for online content. To achieve



What is the CoNEXT'19 paper about?



1. The mapping problem: Still a valid and important issue



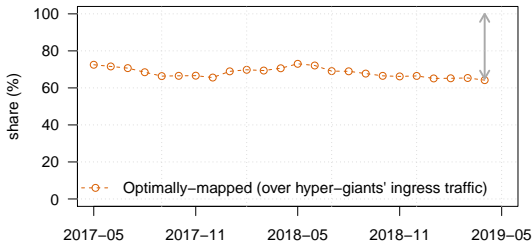
1. The mapping problem: Still a valid and important issue
2. From PaDIS to FlowDirector: Changes to the initial system



1. The mapping problem: Still a valid and important issue
2. From PaDIS to FlowDirector: Changes to the initial system
3. FlowDirector deployment: 2 years of operational experience



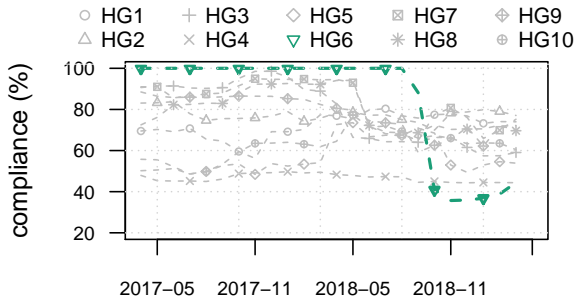
USER-TO-SERVER MAPPING PROBLEM



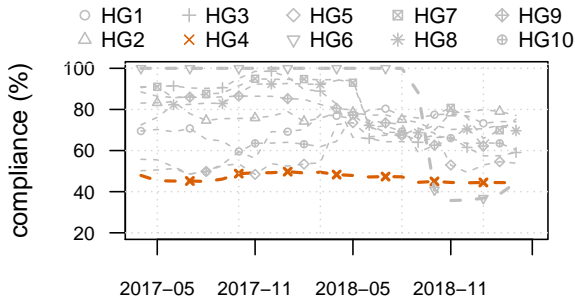
Optimally-mapped: Ingress via the PoP with lowest cost ³

- $\approx 35\%$ of traffic is not optimally-mapped
- steady negative trend

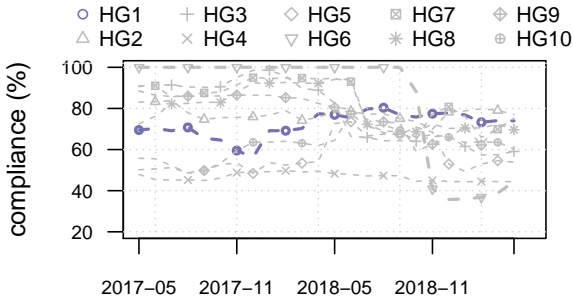
³Combination of number of hops and their distances with each other



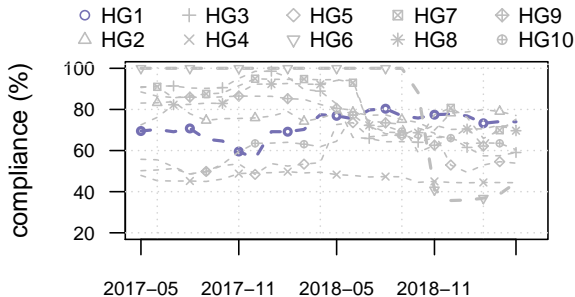
Challenges: Peering at a new location is difficult...



Incentives: Sometimes there are no direct incentives...

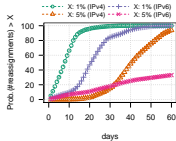
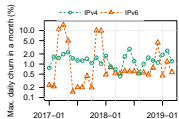
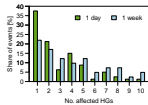
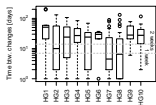
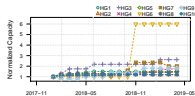
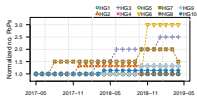


Accuracy: Some do actually try and get good results...

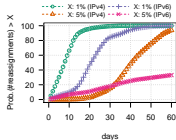
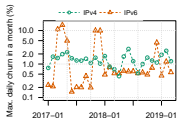
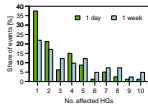
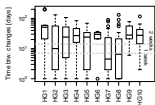
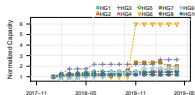
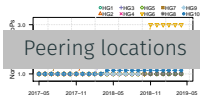


Why is getting 100% compliance difficult?

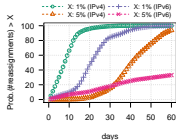
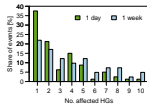
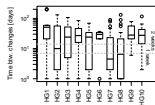
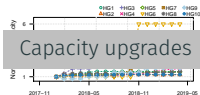
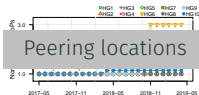
USER-TO-SERVER MAPPING IS A DIFFICULT PROBLEM



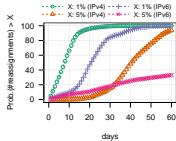
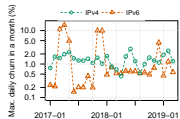
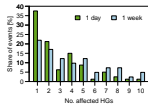
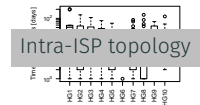
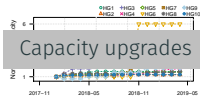
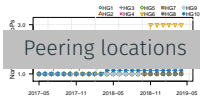
USER-TO-SERVER MAPPING IS A DIFFICULT PROBLEM



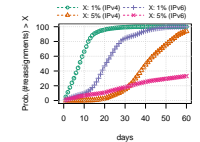
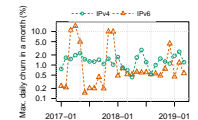
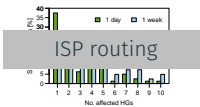
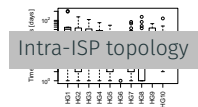
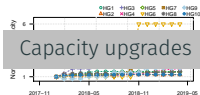
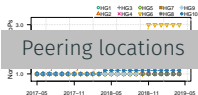
USER-TO-SERVER MAPPING IS A DIFFICULT PROBLEM



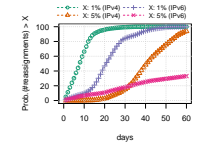
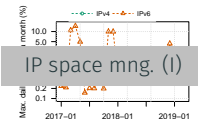
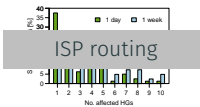
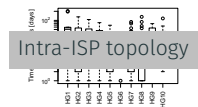
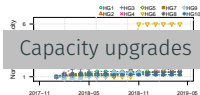
USER-TO-SERVER MAPPING IS A DIFFICULT PROBLEM



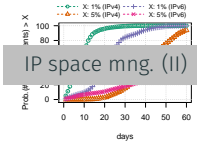
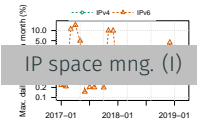
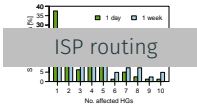
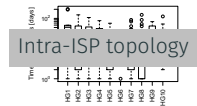
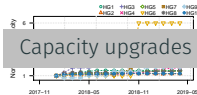
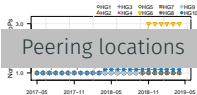
USER-TO-SERVER MAPPING IS A DIFFICULT PROBLEM



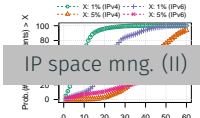
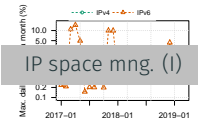
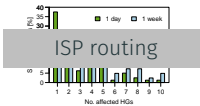
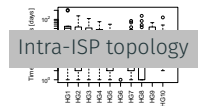
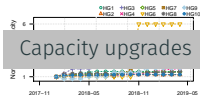
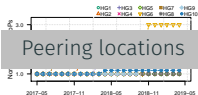
USER-TO-SERVER MAPPING IS A DIFFICULT PROBLEM



USER-TO-SERVER MAPPING IS A DIFFICULT PROBLEM



USER-TO-SERVER MAPPING IS A DIFFICULT PROBLEM



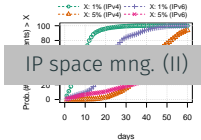
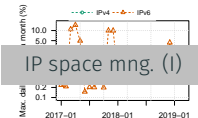
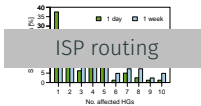
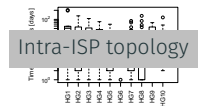
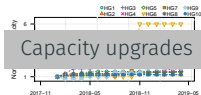
Unknown factors:

- Server loads
- Maintenance
- Content availability

Other:

- Cross traffic

USER-TO-SERVER MAPPING IS A DIFFICULT PROBLEM



Unknown factors:

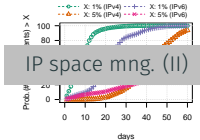
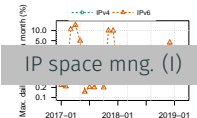
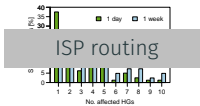
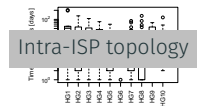
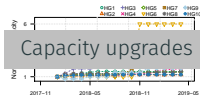
- Server loads
- Maintenance
- Content availability

Other:

- Cross traffic

More details in the paper

USER-TO-SERVER MAPPING IS A DIFFICULT PROBLEM



Unknown factors:

- Server loads
- Maintenance
- Content availability

Other:

- Cross traffic

More details in the paper

Lack of visibility: Collaboration to the rescue!



FROM PADIS TO FLOWDIRECTOR



1. Collects data to determine the state of the ISP's network
 - 1.1 Determine forwarding path from control plane
 - 1.2 Optional: Inventory and performance data



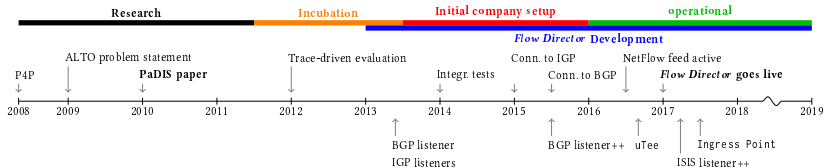
1. **Collects data to determine the state of the ISP's network**
 - 1.1 Determine forwarding path from control plane
 - 1.2 Optional: Inventory and performance data

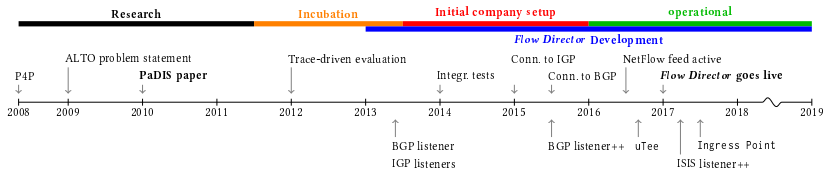
2. **Computes the best ingress location for each customer prefix**
 - 2.1 Ingress-point detection from data plane (server subnets)



1. **Collects data to determine the state of the ISP's network**
 - 1.1 Determine forwarding path from control plane
 - 1.2 Optional: Inventory and performance data
2. **Computes the best ingress location for each customer prefix**
 - 2.1 Ingress-point detection from data plane (server subnets)
3. **Communicates with the cooperating hyper-giant**
 - 3.1 Automated, near real-time via ALTO, out-of-band BGP, etc.

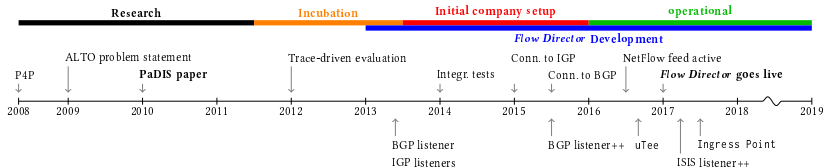
FROM A RESEARCH IDEA TO A PRODUCTION SYSTEM





Components design:

- RFC conforming input
- Customizable output
- Horizontally scalable



Components design:

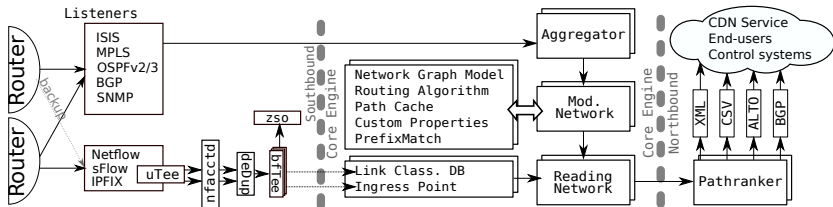
- RFC conforming input
- Customizable output
- Horizontally scalable

Operational requirements:

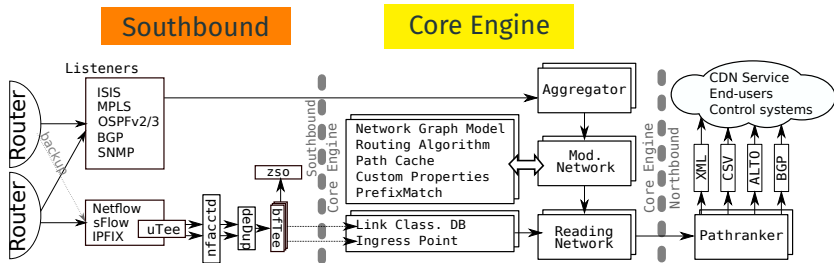
- safe, secure, and redundant IGP
- $\sim 1 \frac{\text{Gbit}}{\text{sec}}$ Netflow
- ~ 600 BGP sessions
- $\sim 60\text{s}$ reaction time



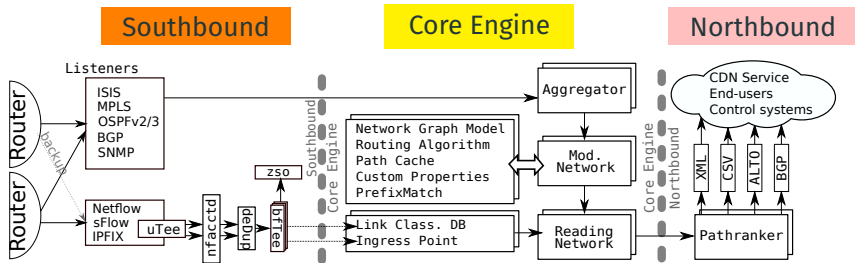
Southbound



Details in the paper...



Details in the paper...



Details in the paper...



OPERATIONAL EXPERIENCE

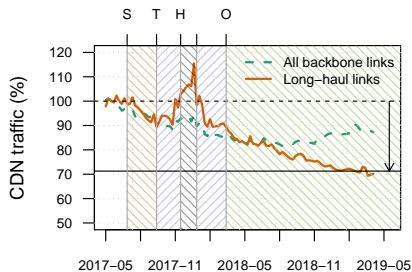


Overview:

- > 10% of the ISP's ingress traffic and multiple ingress PoPs
- KPIs:
 - for the ISP: reduce long-haul traffic
 - for the hyper-giant: reduce latency
- function: combination path length and distance
- FD's suggestion can be ignored
- progressive roll-out



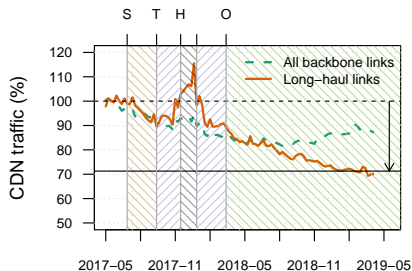
Combined with network planning:
30% reduction long-haul traffic



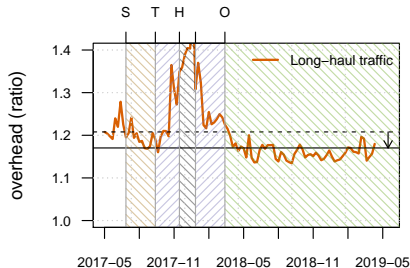
S=Start T=Test H=Hold O=Operational



Combined with network planning:
30% reduction long-haul traffic



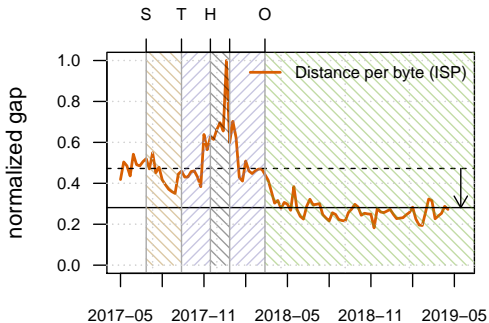
Better mapping:
15% reduction traffic overhead



S=Start T=Test H=Hold O=Operational

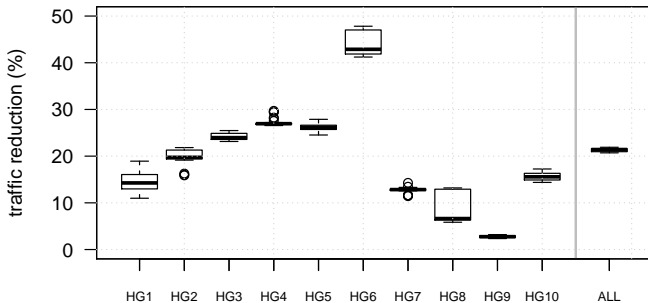


Distance as a proxy for latency:
40% reduction





Upper bounds for long-haul traffic reduction:
20% reduction





Key takeaways:

1. Opportunity to operate networks more efficiently
2. We enabled the first automated hypergiant-ISP collaboration
3. Lots of engineering and diplomacy involved
4. It works!

Next steps:

1. Different optimization functions
2. Federated FlowDirector (multi-ISP collaboration)



Thank you for your attention! Questions?