Steering Hyper-Giants' Traffic at Scale

Enric Pujol[1] I. Poese[1] J. Zerwas[2] G. Smaragdakis[3] A. Feldmann[4]

BENOCS[1] TU München[2] TU Berlin[3] Max Planck Inst. Informatics[4]

Hyper-giants on the Internet

What are *hyper-giants*? [1][2]

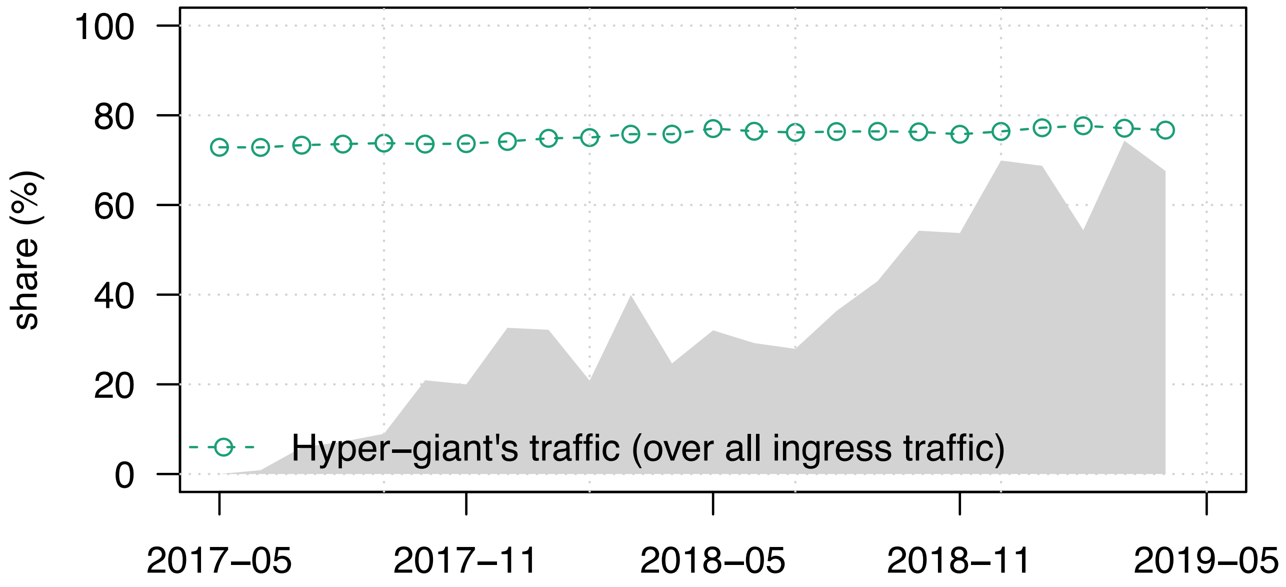
- Large networks providing services
- Global infrastructure
- Generate enormous amounts of traffic

Some of them...

Akamai
Google
Limelight NETWORKS
facebook
amazon
NETFLIX
CLOUDFLARE
edgecast

[1] Labovitz et. al. “Internet Inter-Domain Traffic” in SIGCOMM’10
[2] Böttger et. al. “Looking for hypergiants in peeringDB.” ACM CCR 48.3

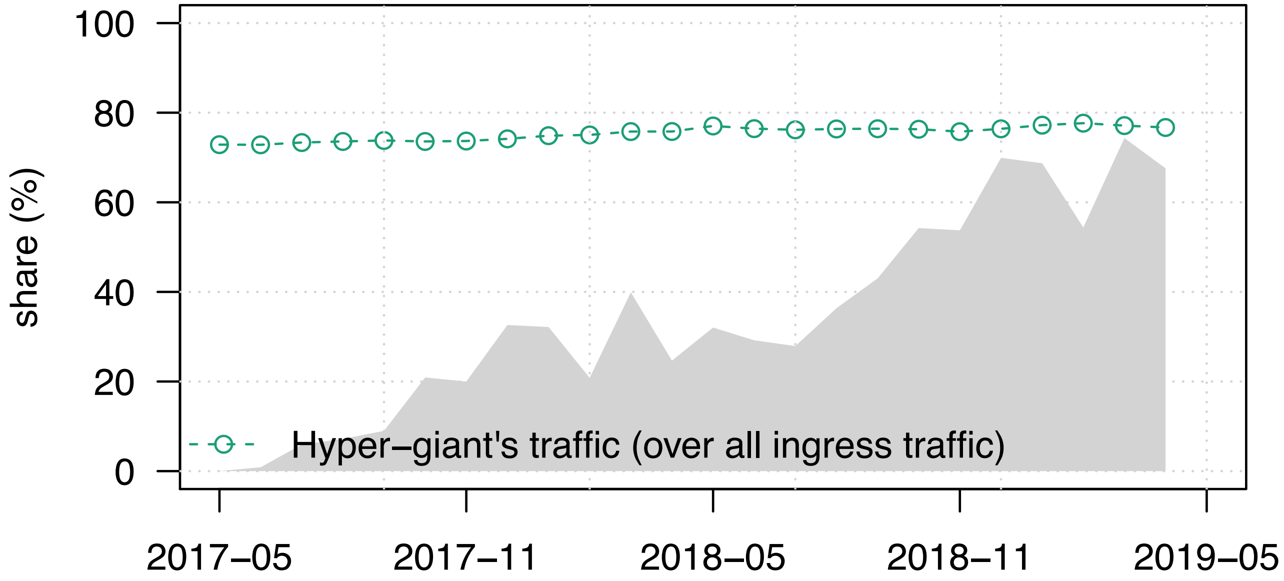
Hyper-giants' traffic



A large ISP's perspective:

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

Hyper-giants' traffic



A large ISP's perspective:

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

Overall ingress traffic:

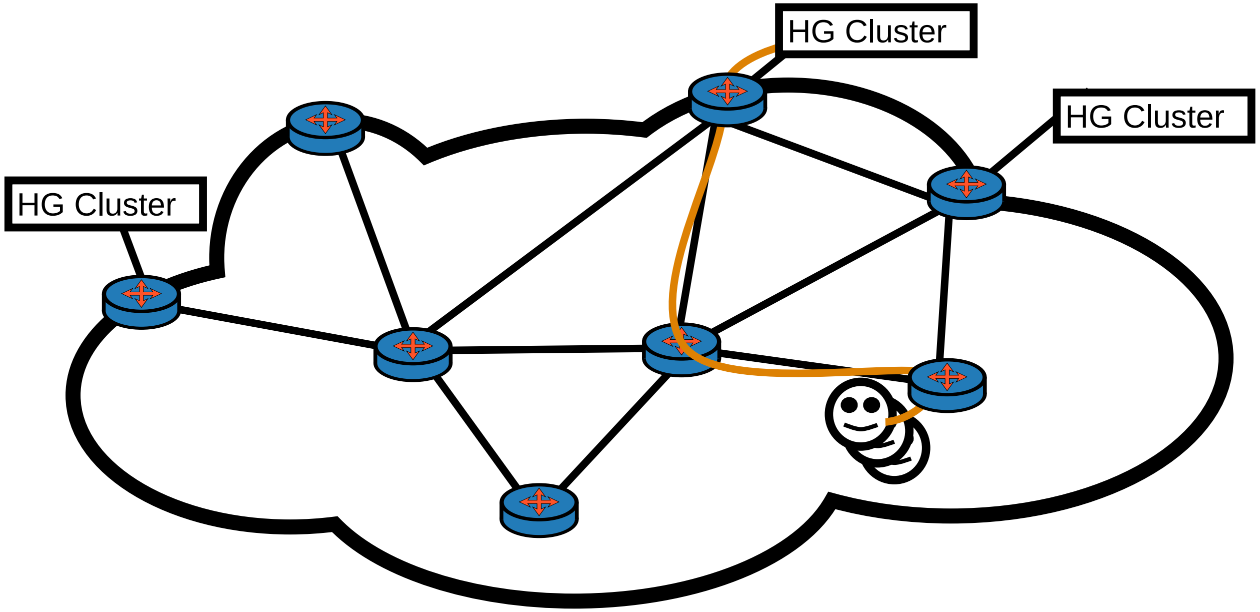
- $\sim 30\,\%$ growth per annum

Top 10 hyper-giants:

- $\sim 75\,\%$ share

Steering hyper-giants' traffic: Motivation

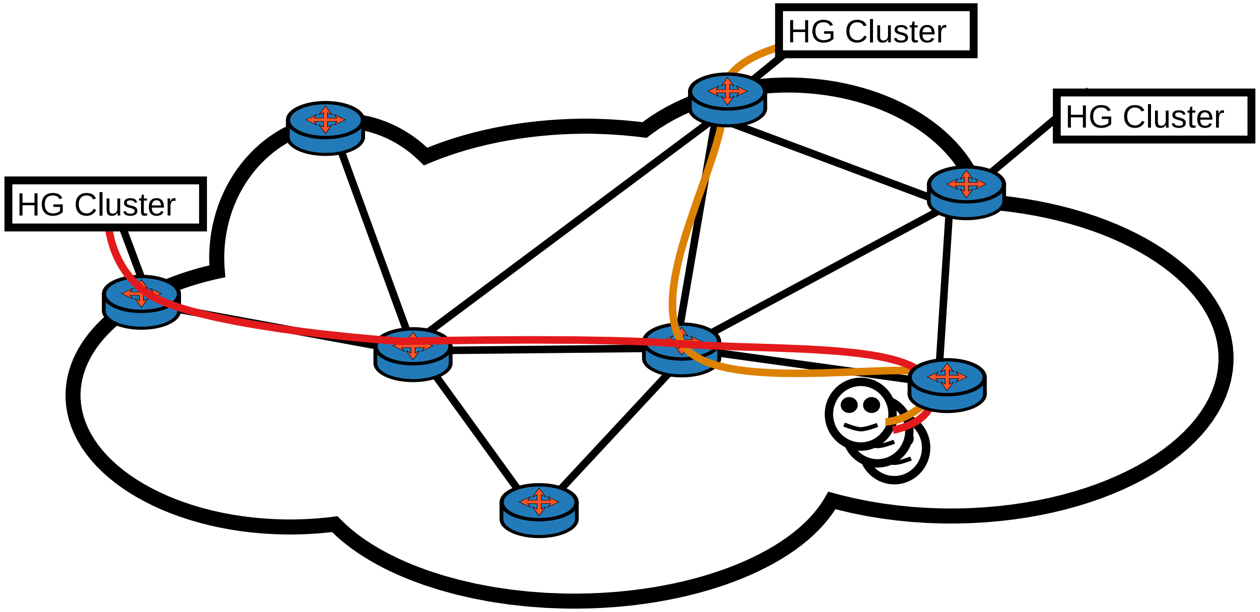
Toy example



Baseline: 2 bytes in the backbone per ingress byte

Steering hyper-giants' traffic: Motivation

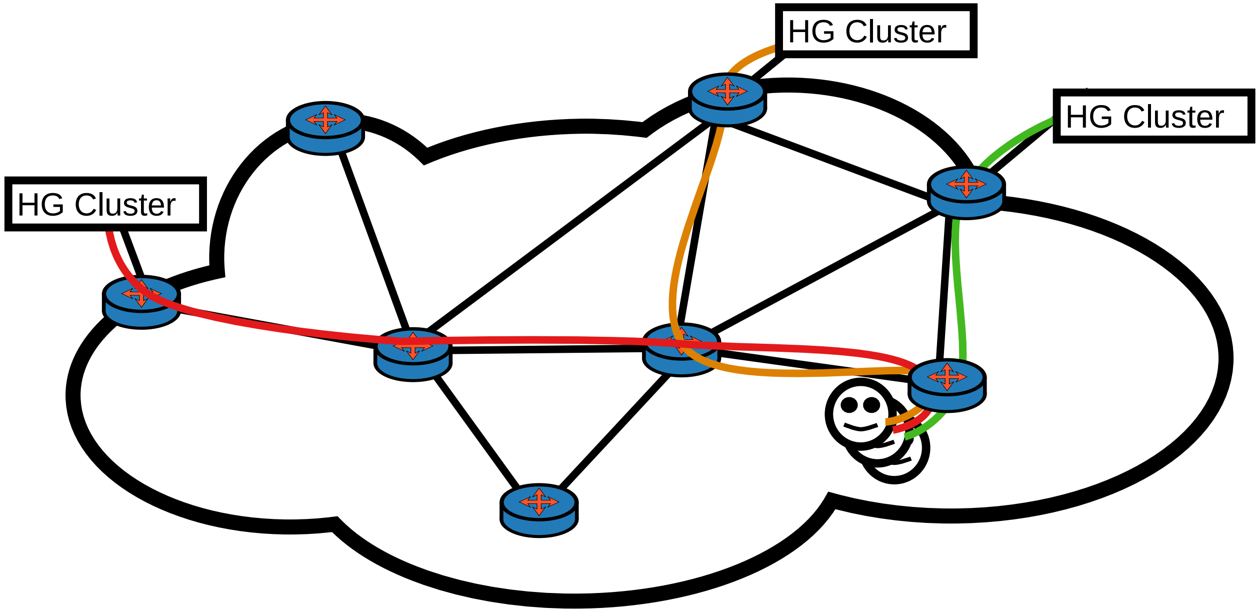
Toy example



“Bad” mapping= higher costs and incr. latency

Steering hyper-giants' traffic: Motivation

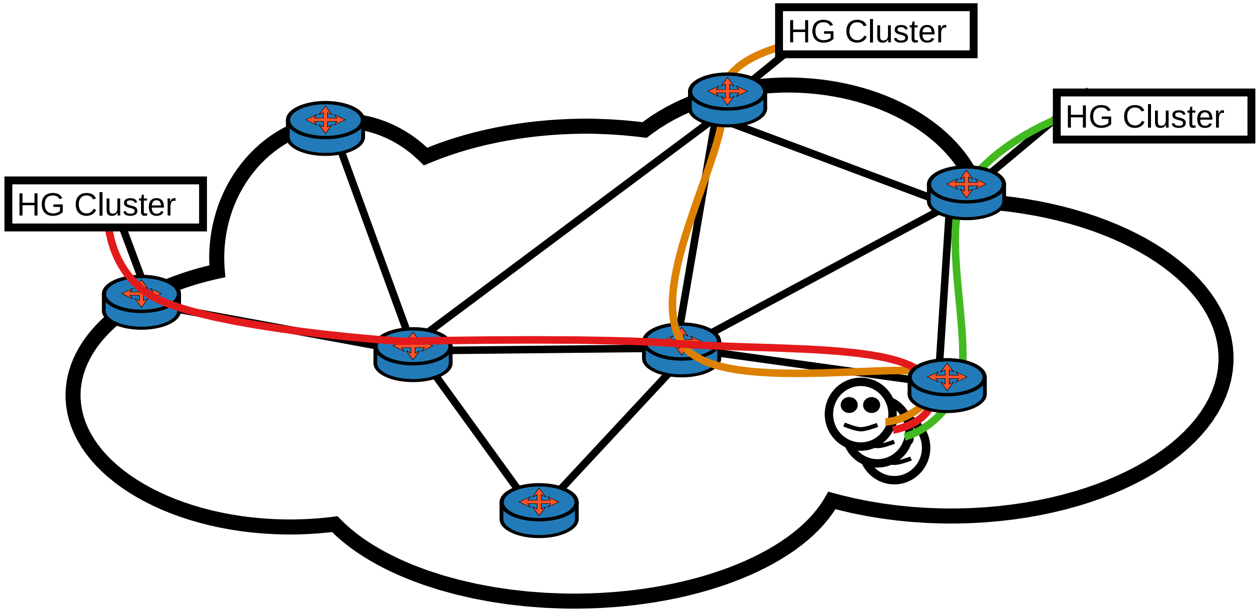
Toy example



"**Better" mapping**= 50% reduction

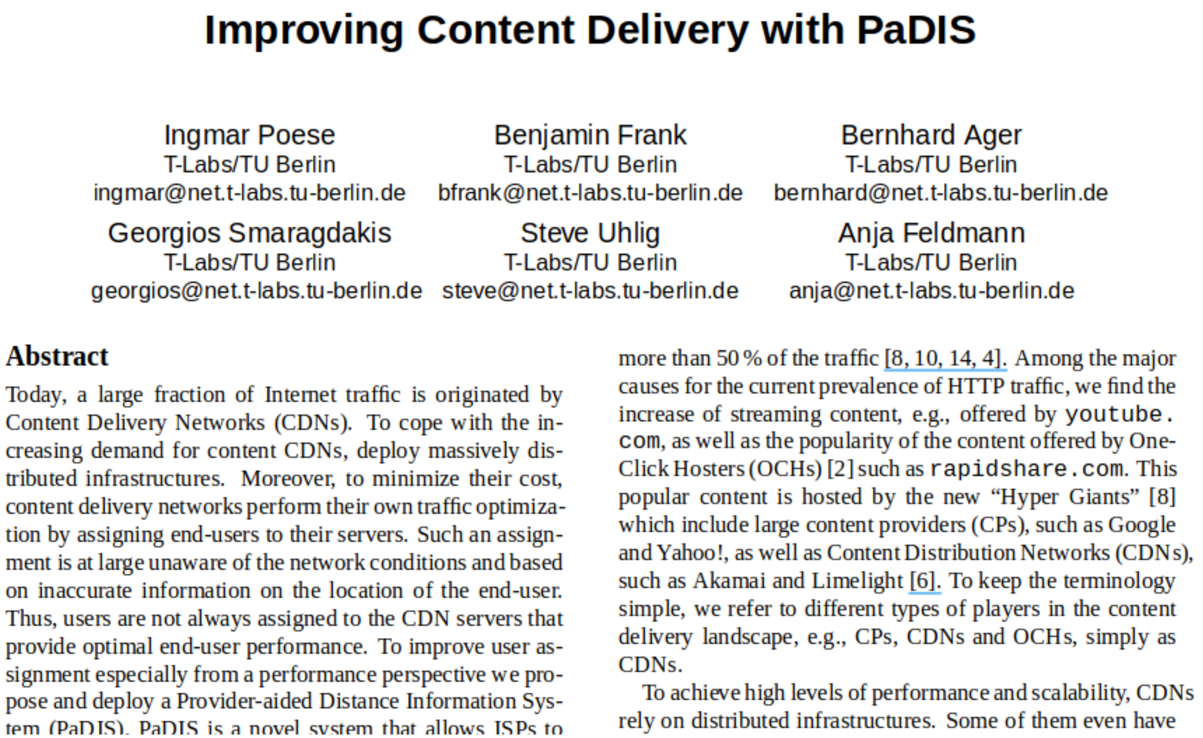
Steering hyper-giants' traffic: Motivation

Toy example



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

CDN-ISP collaboration already proposed in 2010



Improving Content Delivery with PaDIS

Ingmar Poese
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

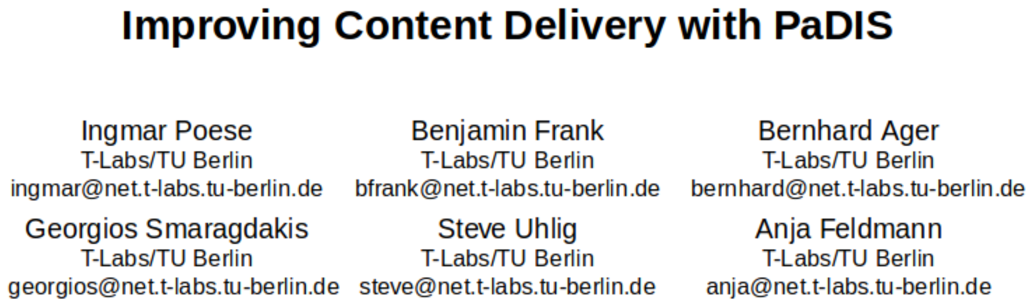
Abstract

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

CDN-ISP collaboration already proposed in 2010



Improving Content Delivery with PaDIS

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

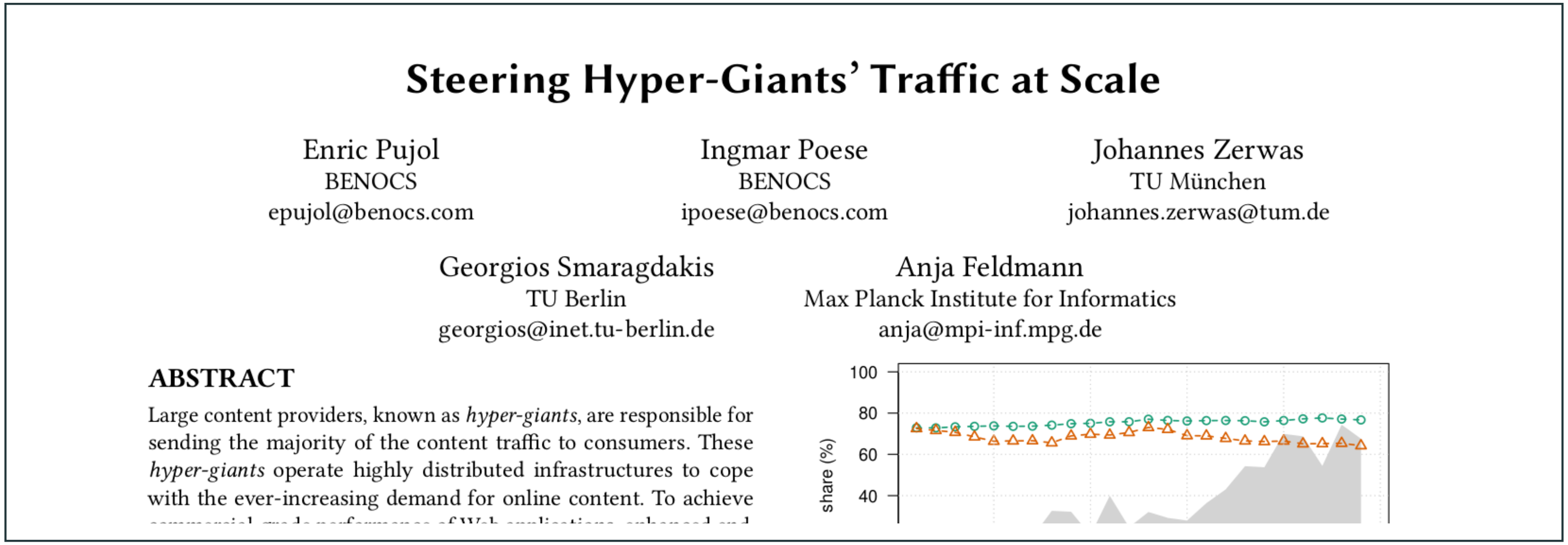
Benjamin Frank
T-Labs/TU Berlin
bfrank@net.t-labs.tu-berlin.de

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 Poese
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 commercial-grade performance of Web applications, enhanced and

What is the CoNEXT'19 paper about?

Our contributions

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

Our contributions

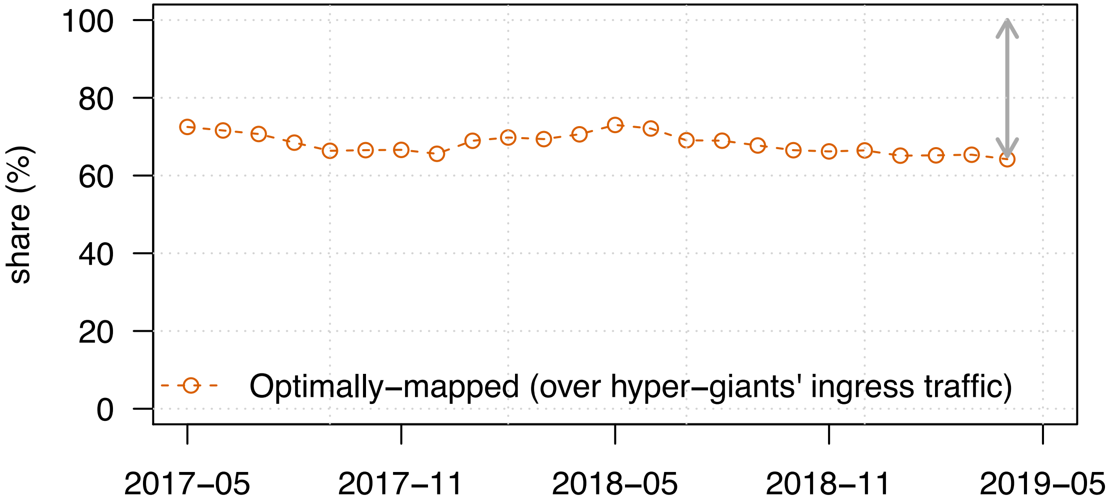
1. The mapping problem: Still a valid and important issue
2. From PaDIS to FlowDirector: Changes to the initial system

Our contributions

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

Overall fraction of optimally-mapped traffic

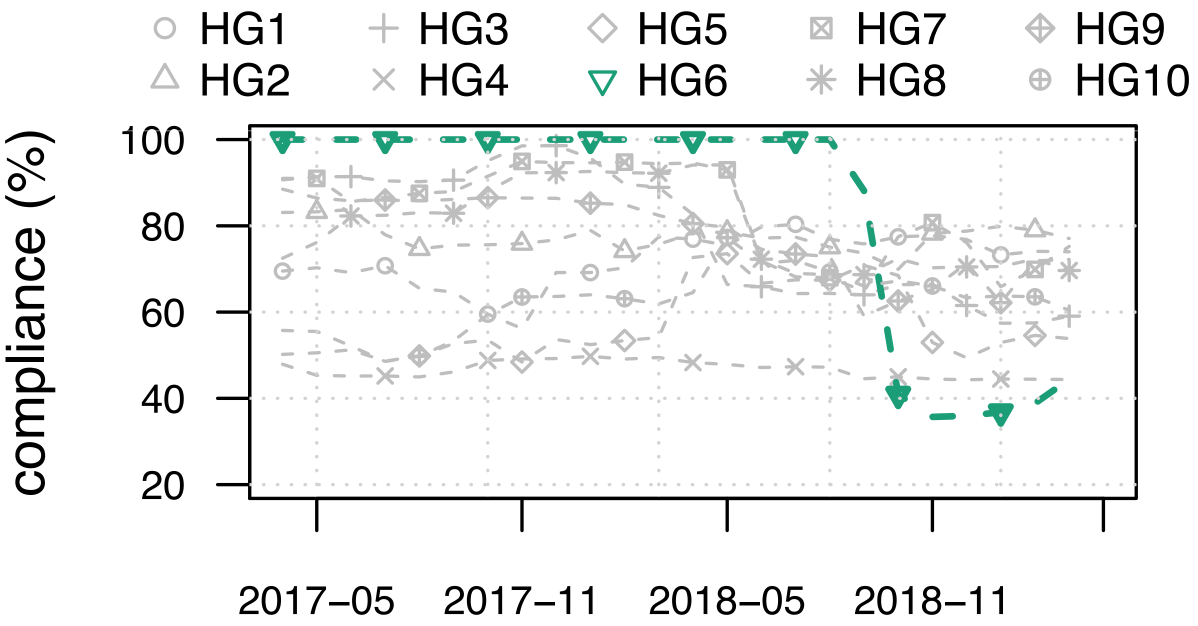


Optimally-mapped: Ingress via the PoP with lowest cost [3]

- ≈35% of traffic is not optimally-mapped
- steady negative trend

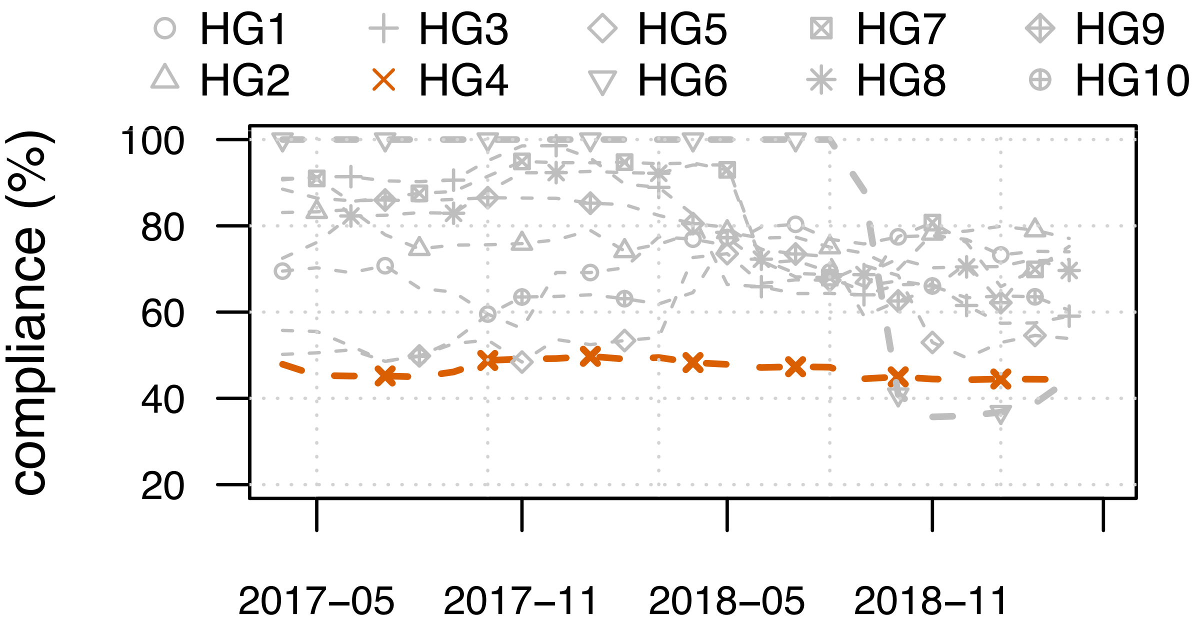
[3] Combination of number of hops and their distances with each other

Optimally-mapped traffic per hyper-giant



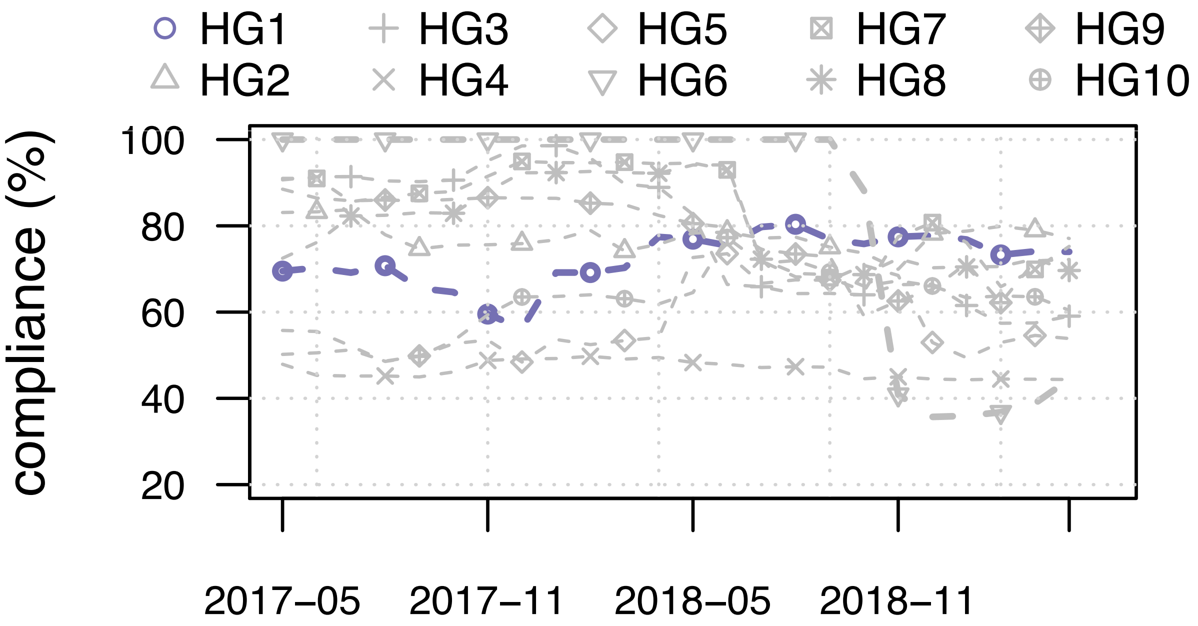
Challenges: Peering at a new location is difficult...

Optimally-mapped traffic per hyper-giant



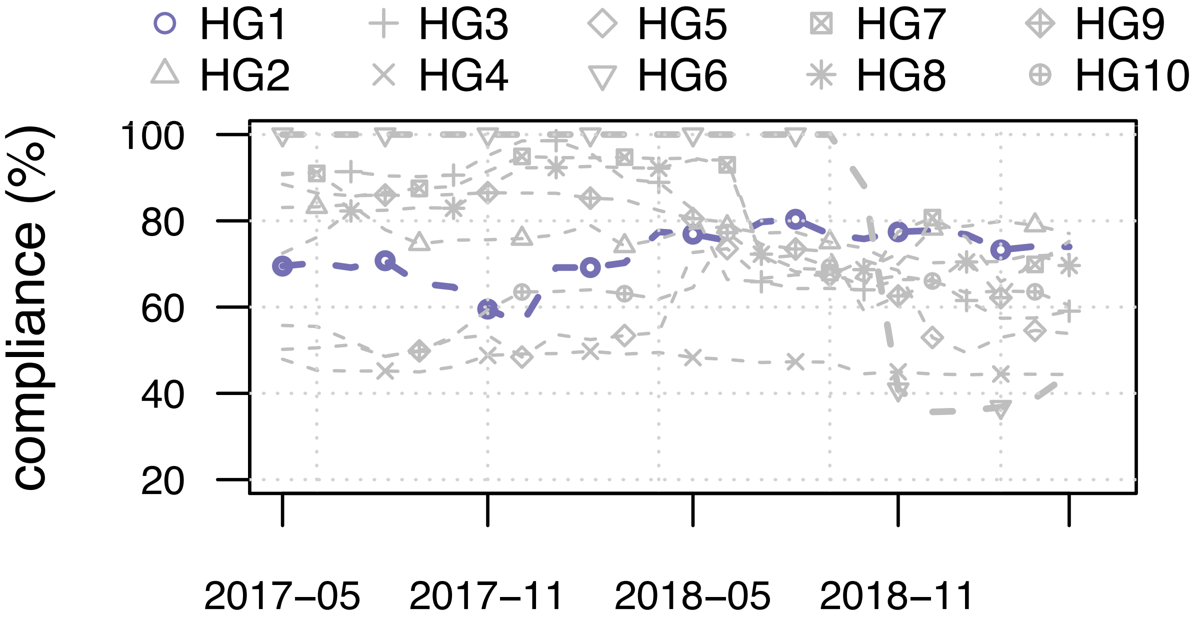
Incentives: Sometimes there are no direct incentives...

Optimally-mapped traffic per hyper-giant



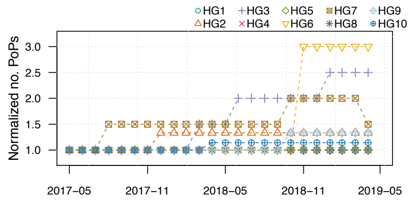
Accuracy: Some do actually try and get good results…

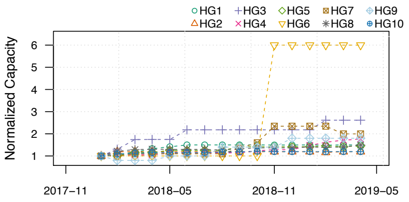
Optimally-mapped traffic per hyper-giant

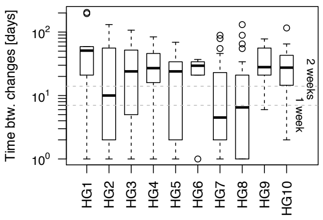


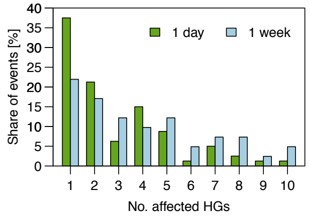
Why is getting 100% compliance difficult?

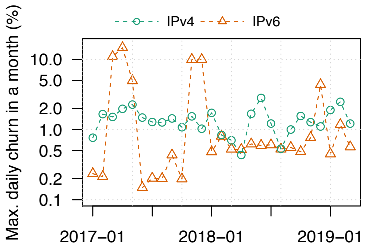
User-to-server mapping is a difficult problem

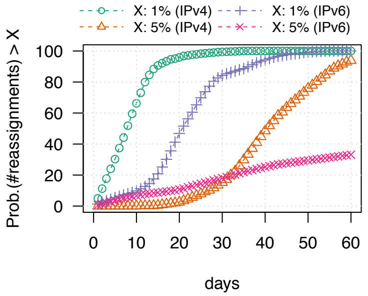




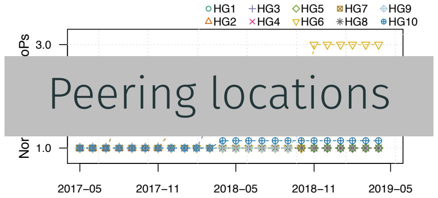


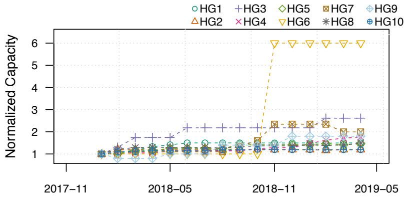


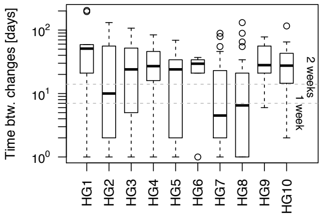


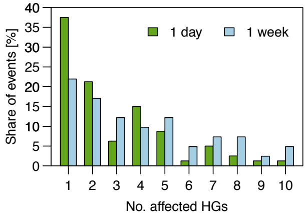


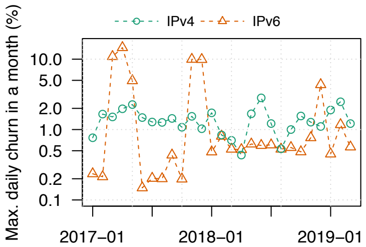
User-to-server mapping is a difficult problem

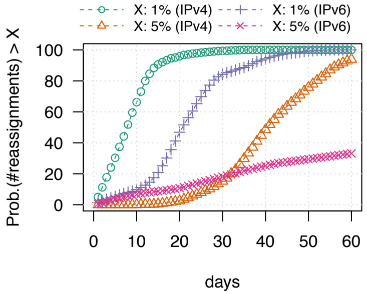

Peering locations



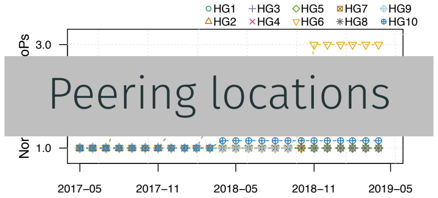




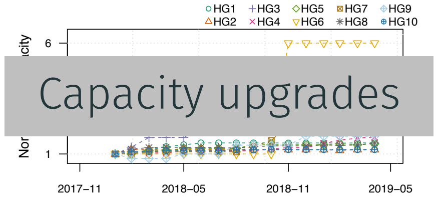




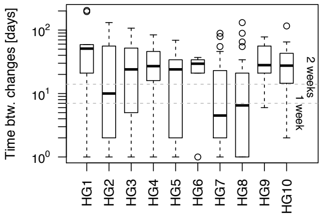
User-to-server mapping is a difficult problem

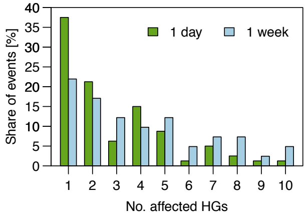


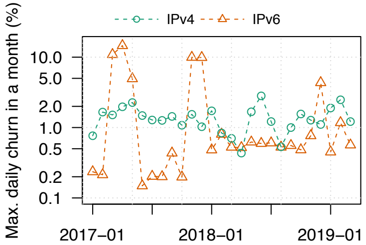
Peering locations

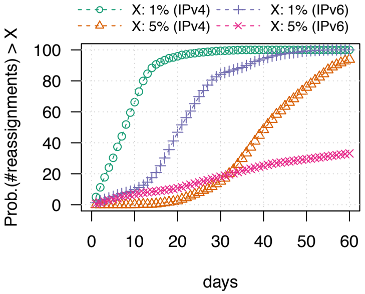


Capacity upgrades

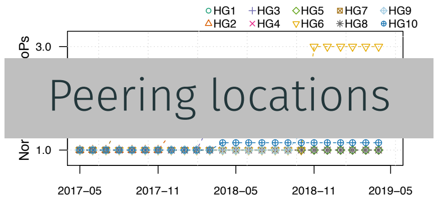




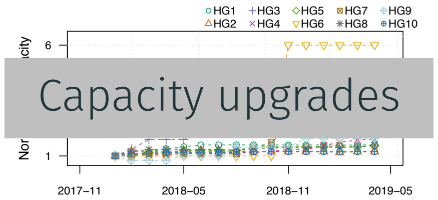




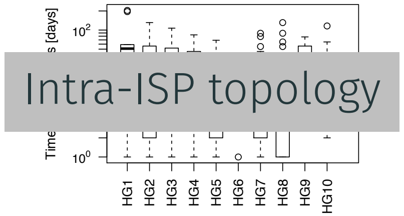
User-to-server mapping is a difficult problem



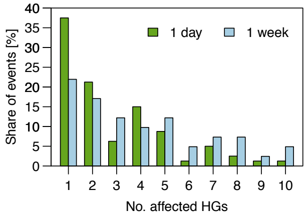
Peering locations

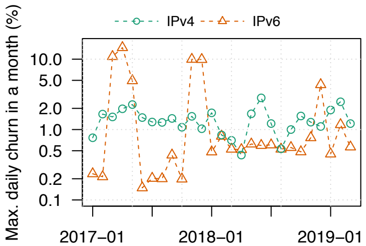


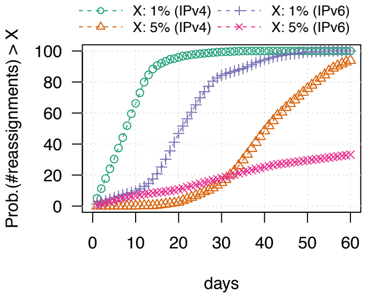
Capacity upgrades



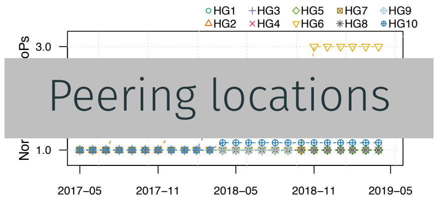
Intra-ISP topology



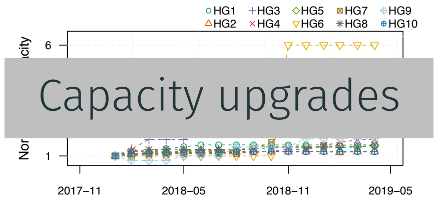




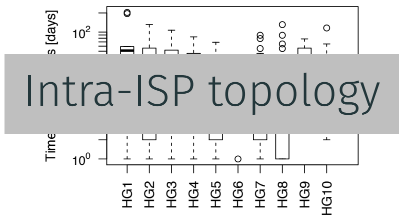
User-to-server mapping is a difficult problem



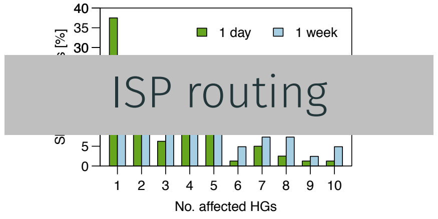
Peering locations



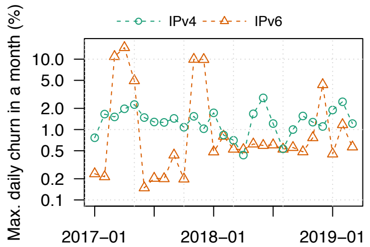
Capacity upgrades

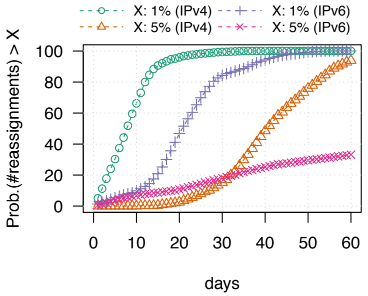


Intra-ISP topology

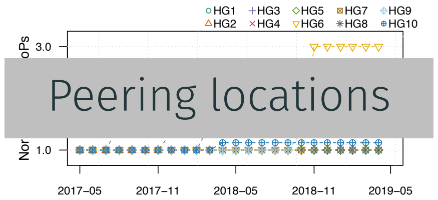


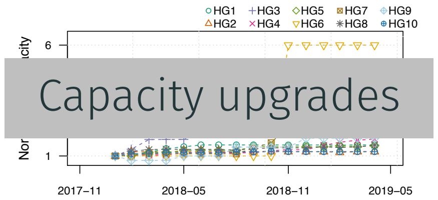
ISP routing

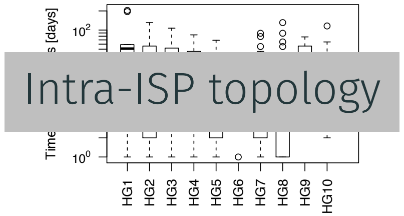


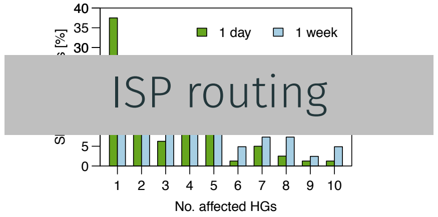


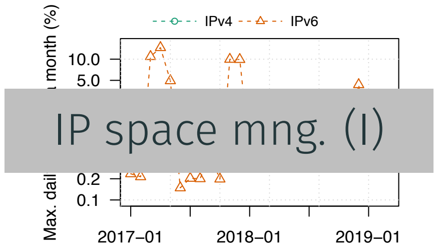
User-to-server mapping is a difficult problem

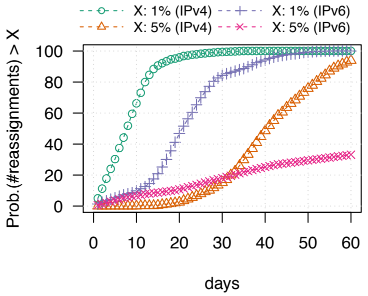

Peering locations


Capacity upgrades

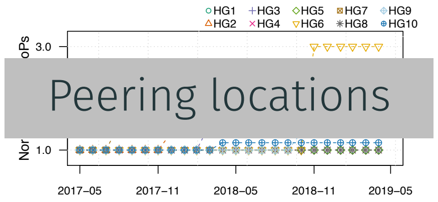

Intra-ISP topology

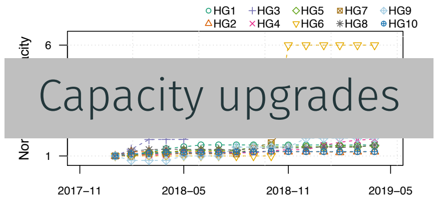

ISP routing

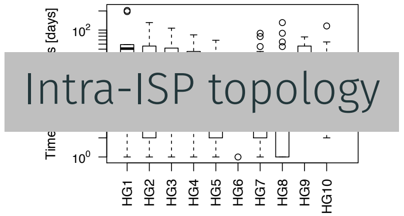

IP space mng. (I)

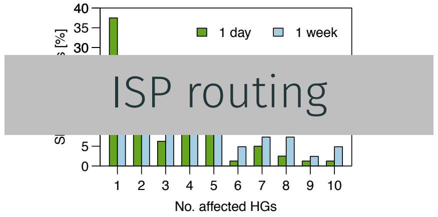


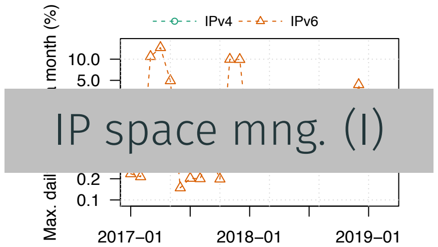
User-to-server mapping is a difficult problem

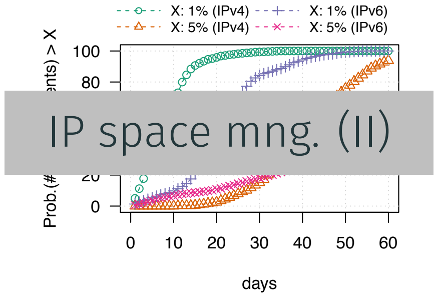

Peering locations


Capacity upgrades

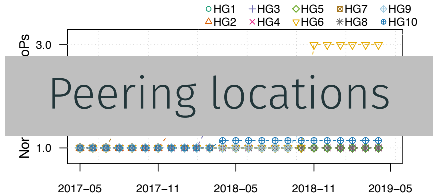

Intra-ISP topology


ISP routing

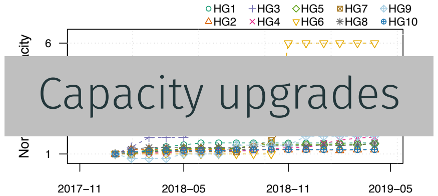

IP space mng. (I)


IP space mng. (II)

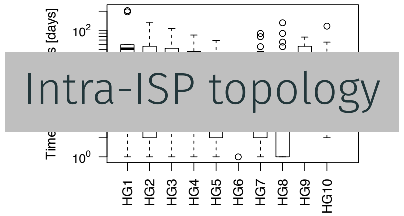
User-to-server mapping is a difficult problem



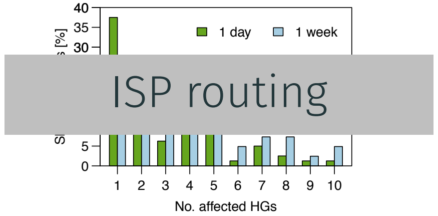
Peering locations



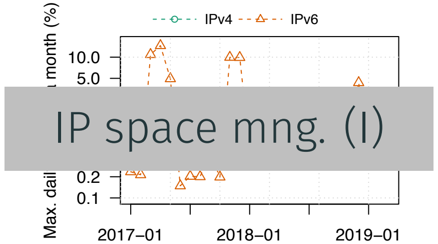
Capacity upgrades



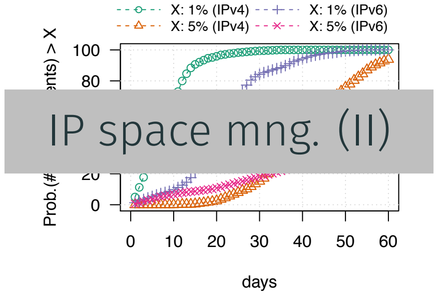
Intra-ISP topology



ISP routing



IP space mng. (I)



IP space mng. (II)

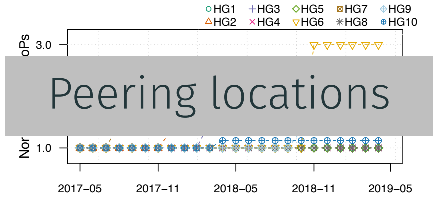
Unknown factors:

- Server loads
- Maintenance
- Content availability

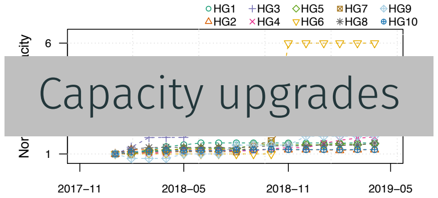
Other:

- Cross traffic

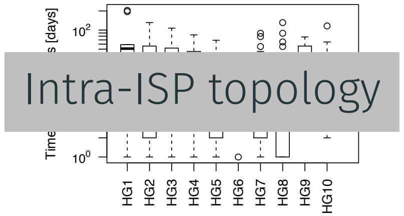
USER-TO-SERVER MAPPING IS A DIFFICULT PROBLEM



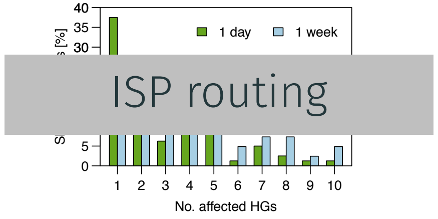
Peering locations



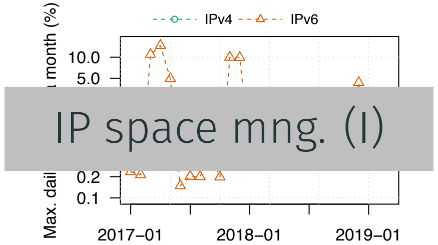
Capacity upgrades



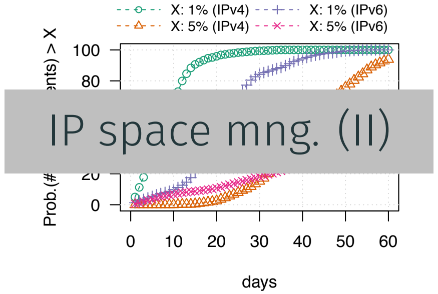
Intra-ISP topology



ISP routing



IP space mng. (I)



IP space mng. (II)

Unknown factors:

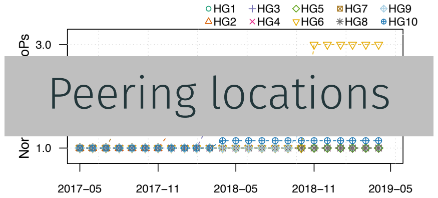
- Server loads
- Maintenance
- Content availability

Other:

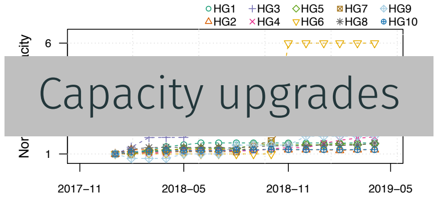
- Cross traffic

More details in the paper

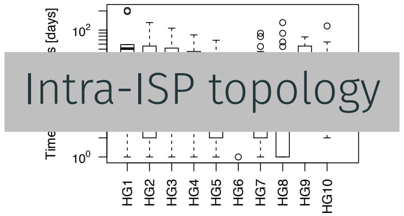
User-to-server mapping is a difficult problem



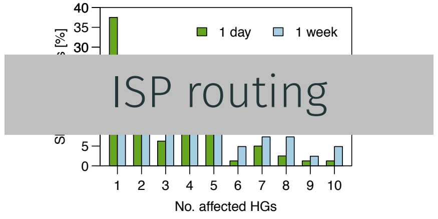
Peering locations



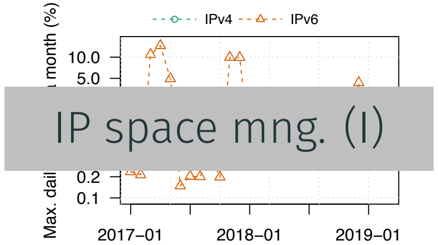
Capacity upgrades



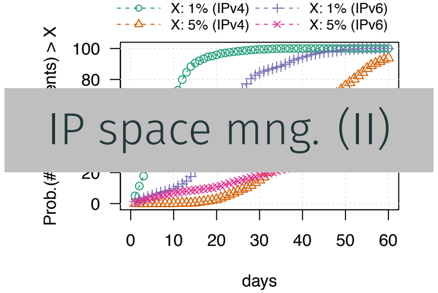
Intra-ISP topology



ISP routing



IP space mng. (I)



IP space mng. (II)

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

The FlowDirector in a nutshell

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

The FlowDirector in a nutshell

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)

The FlowDirector in a nutshell

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

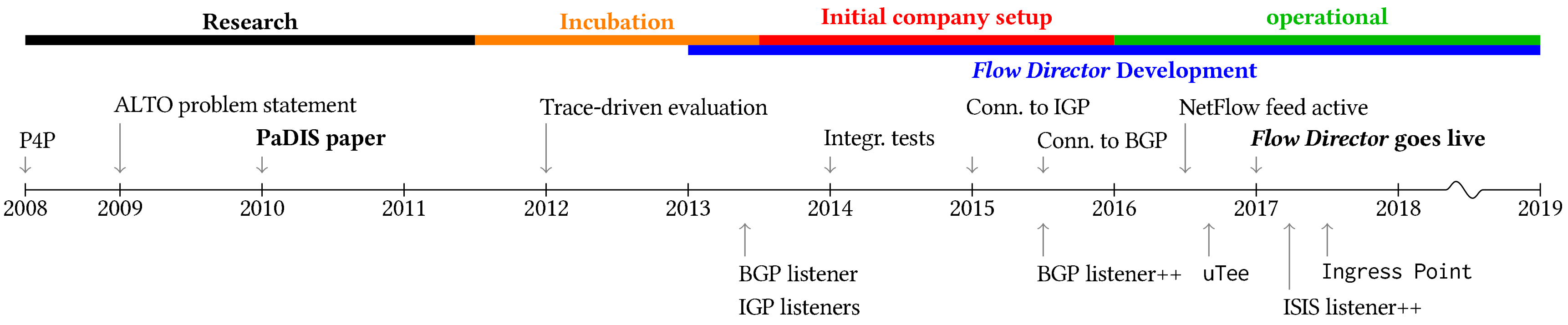
2. **Computes the best ingress location for each customer prefix**

   2.1 Ingress-point detection from data plane (server subnets)

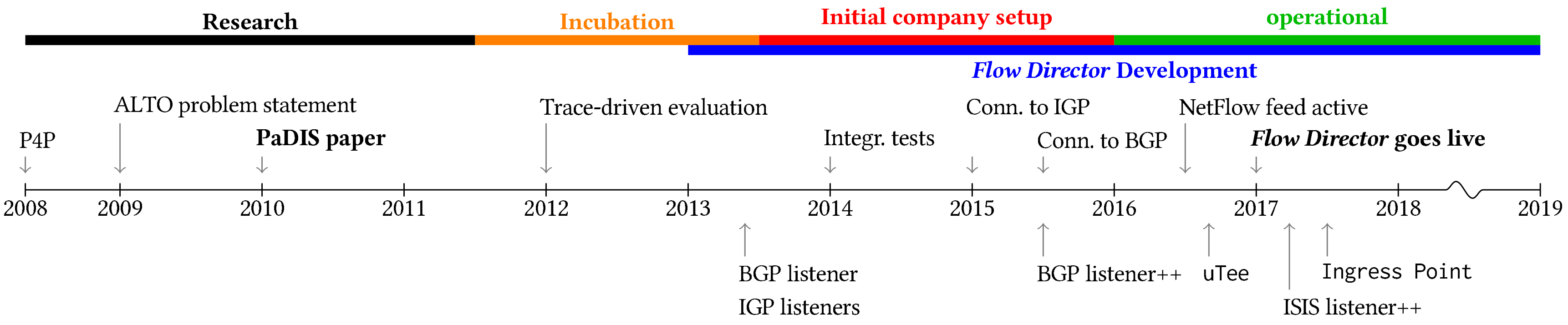
3. **Communicates with the cooperating hyper-giant**

   3.1 Automated, near real-time via ALTO, out-of-band BGP, etc.

From a research idea to a production system



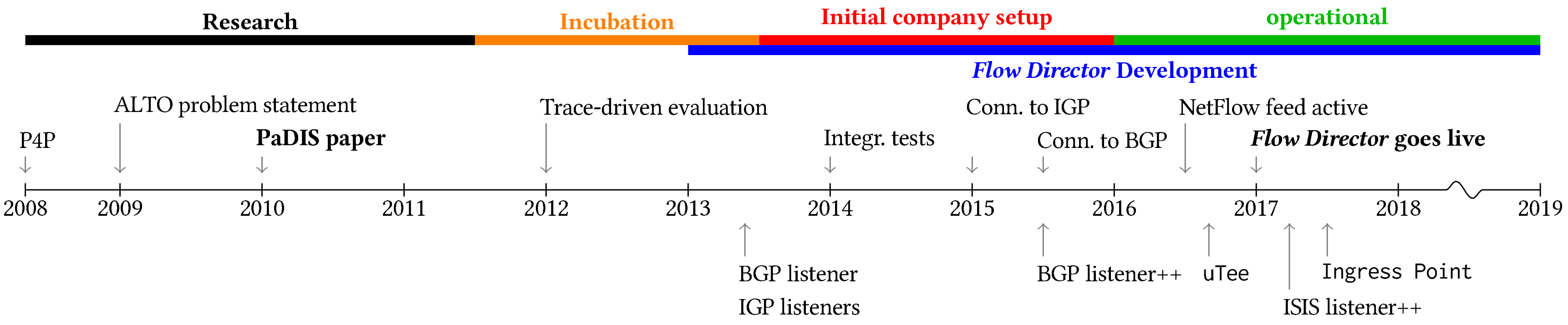
From a research idea to a production system



Components design:

- RFC conforming input
- Customizable output
- Horizontally scalable

From a research idea to a production system



Components design:

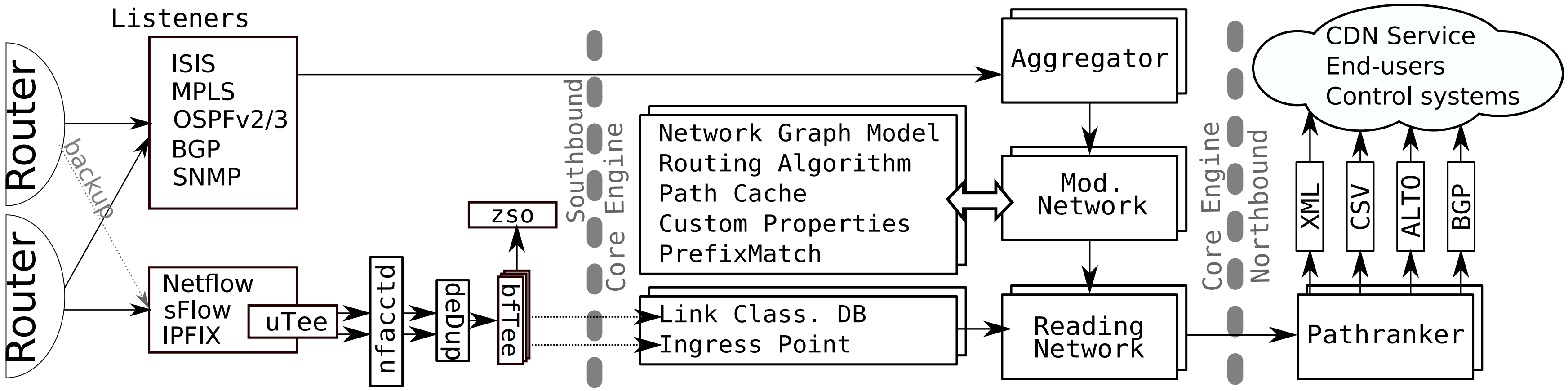
- RFC conforming input
- Customizable output
- Horizontally scalable

Operational requirements:

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

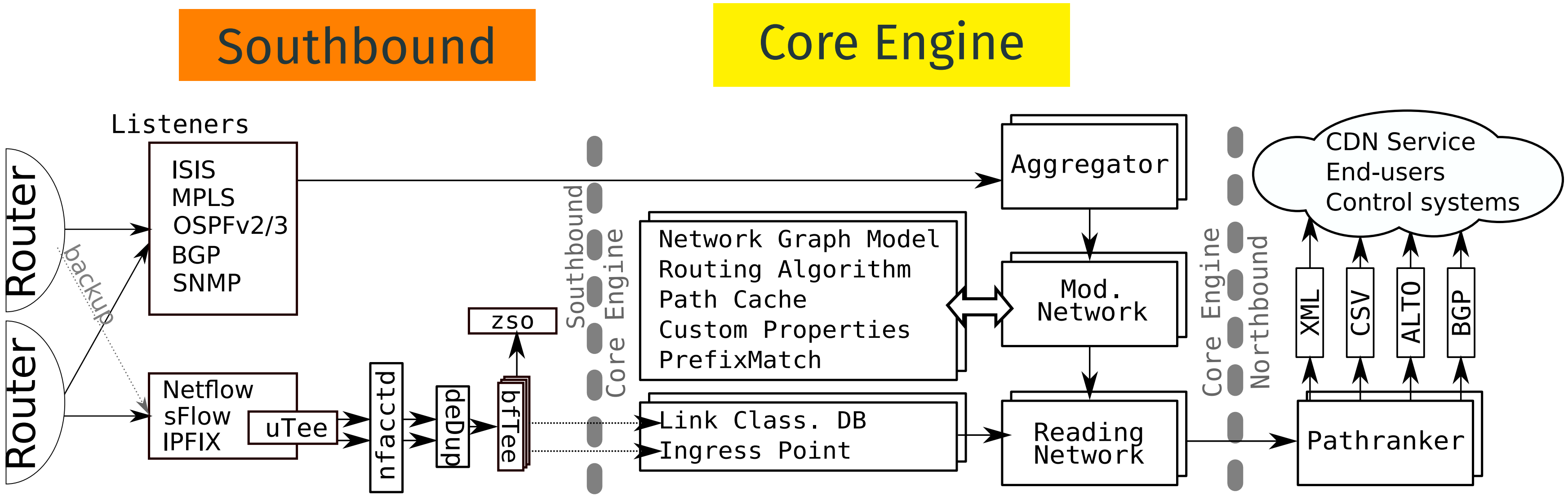
Architecture as of 2019

Southbound



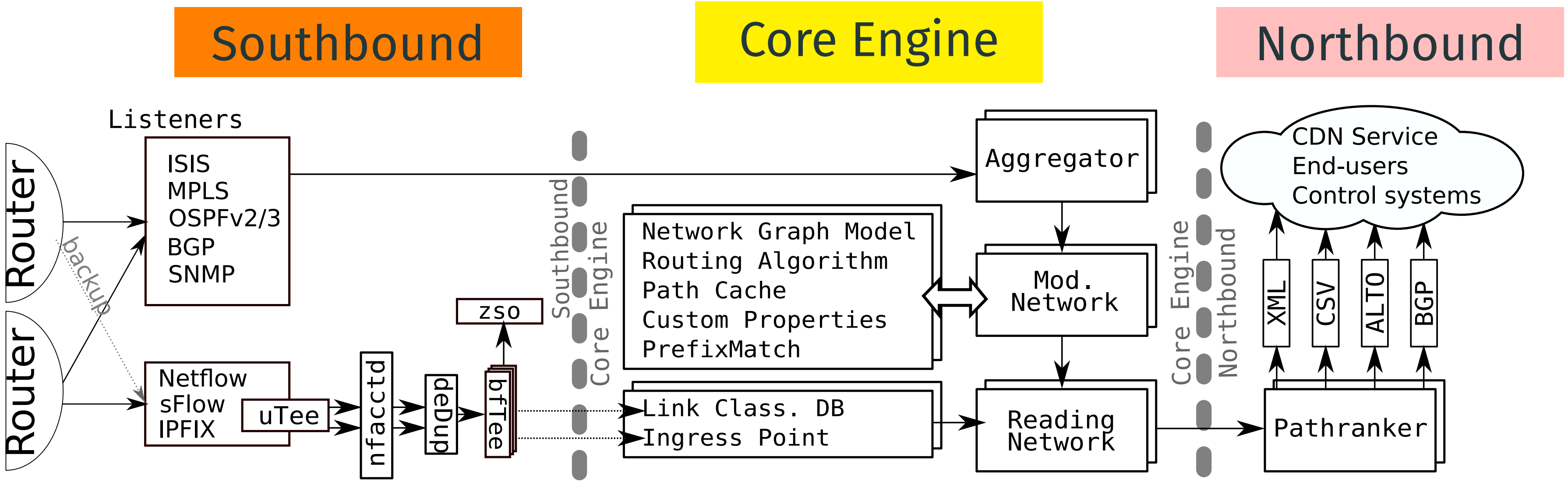
Details in the paper...

ARCHITECTURE AS OF 2019



Details in the paper...

Architecture as of 2019



Details in the paper...

Operational experience

2-YEARS EXPERIENCE WITH ONE COOPERATING HYPER-GIANT

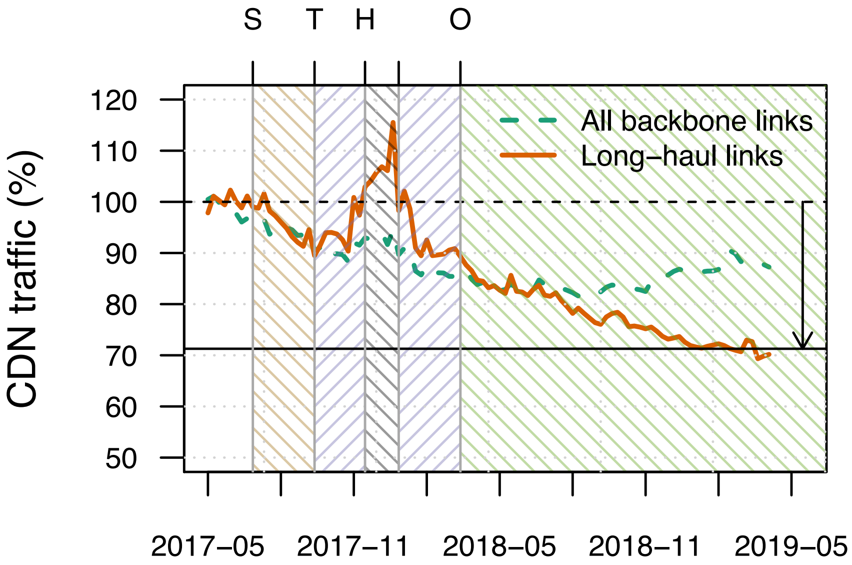
Overview:

- $>$ 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

Benefits for the ISP

Combined with network planning:

30% reduction long-haul traffic

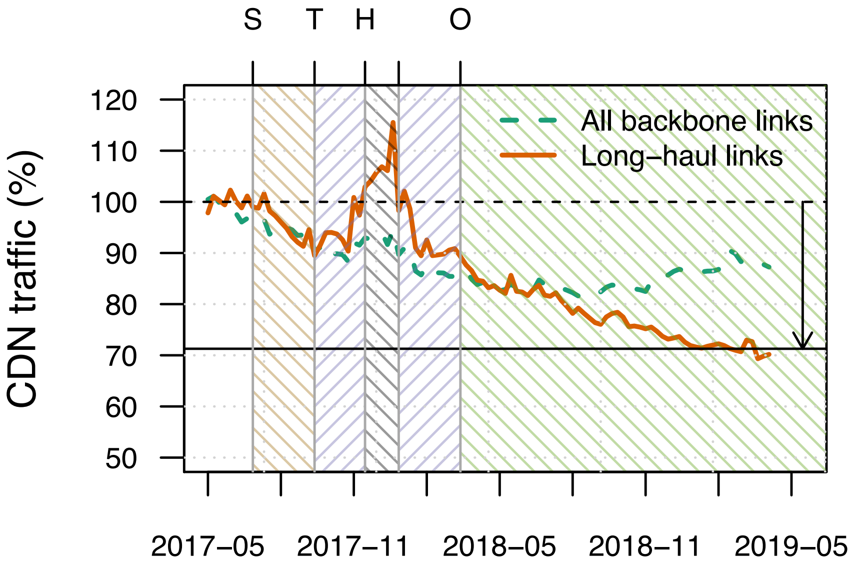


S=Start T=Test H=Hold O=Operational

Benefits for the ISP

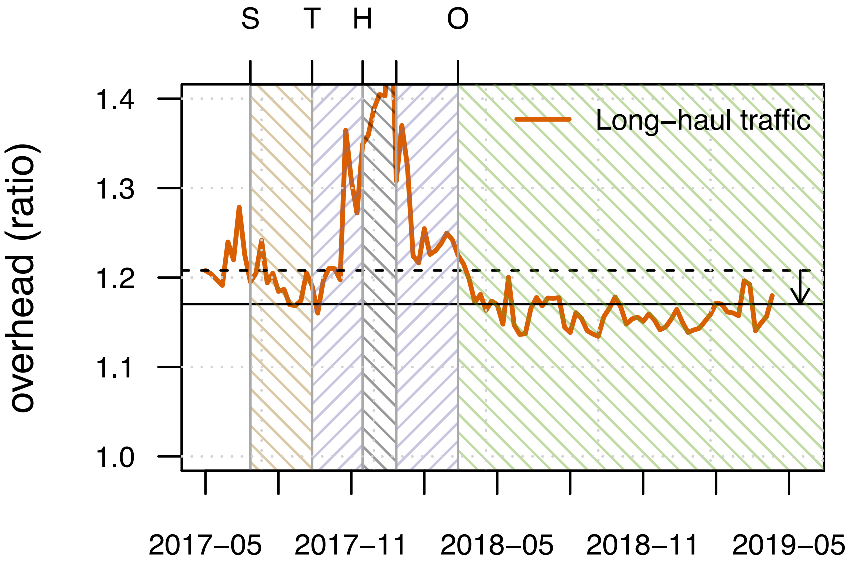
Combined with network planning:

30% reduction long-haul traffic



Better mapping:

15% reduction traffic overhead

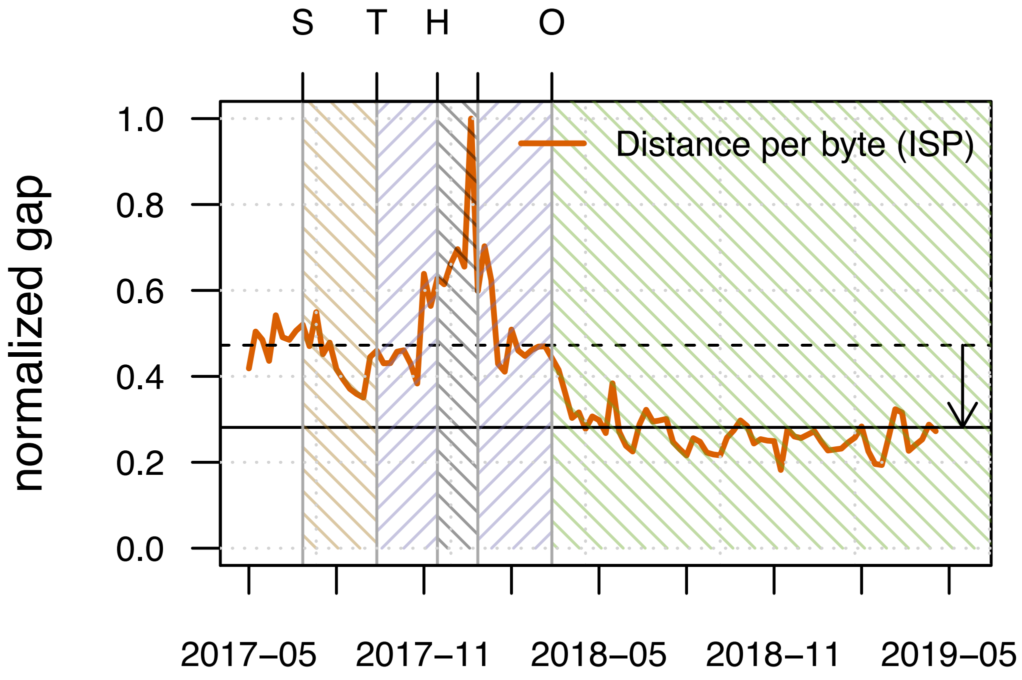


S=Start T=Test H=Hold O=Operational

Benefits for the hyper-giant

Distance as a proxy for latency:

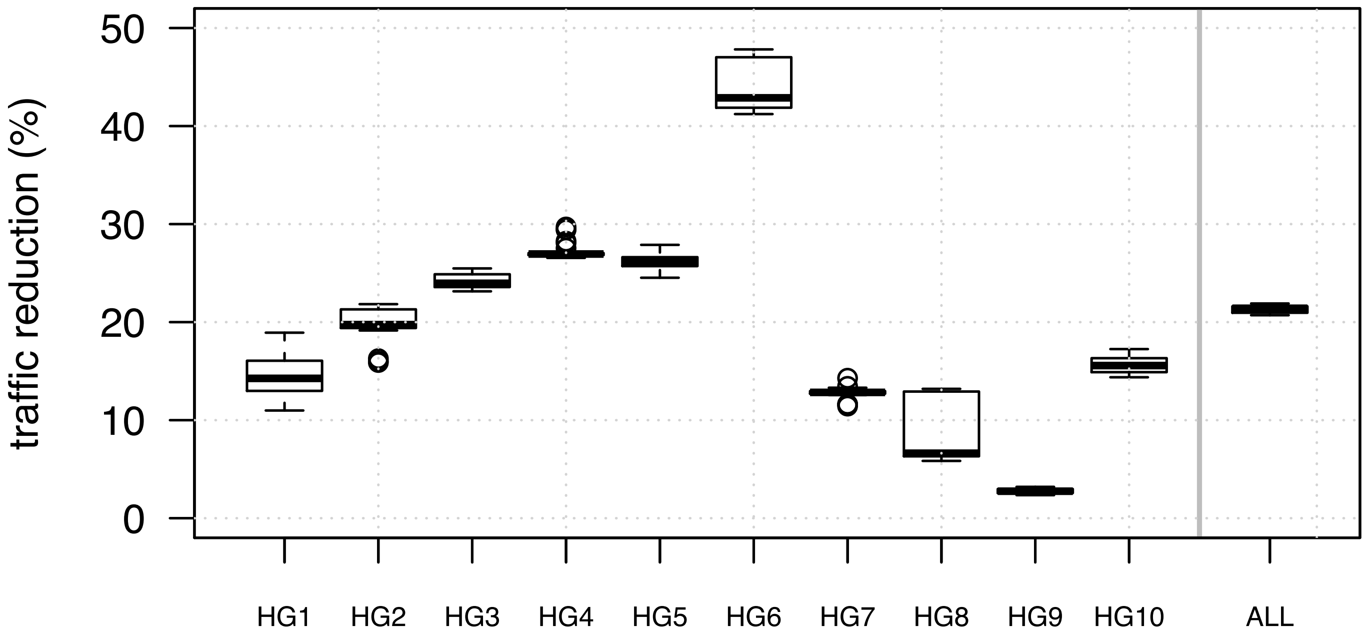
40% reduction



What-if analysis: Hyper-giant on-boarding

Upper bounds for long-haul traffic reduction:

20% reduction



Conclusion and future work

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)

Questions & feedback

Thank you for your attention! Questions?