Deployment experience update

- FlowDirector deployment
- ALTO experiences with ISP-CDN collaboration

Danny Lachos

dlachos@benocs.com

BENOCS GmbH
Reuchlinstraße 10
10553 Berlin, Germany

November 10^{th}, 2021

FlowDirector[1] [2] [3]

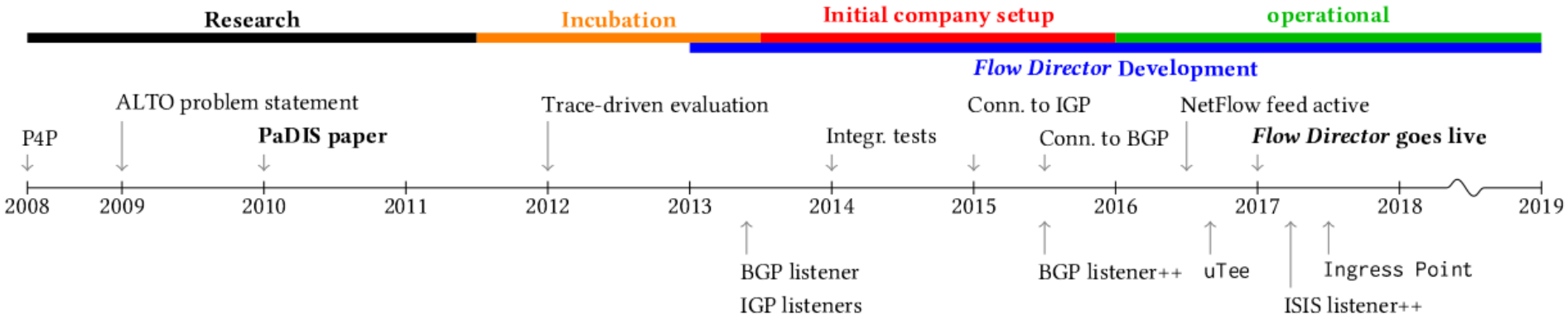
CDN-ISP collaboration system

[1] Source: Steering Hyper-Giants' Traffic at Scale (ACM CoNEXT 2019)
[2] Source: https://irtf.org/anrp/IETF108-ANRP-Poese.pdf
[3] CoNEXT 2019 Best paper award and IETF/IRTF Applied Networking Research Prize 2020

The FlowDirector in a nutshell

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

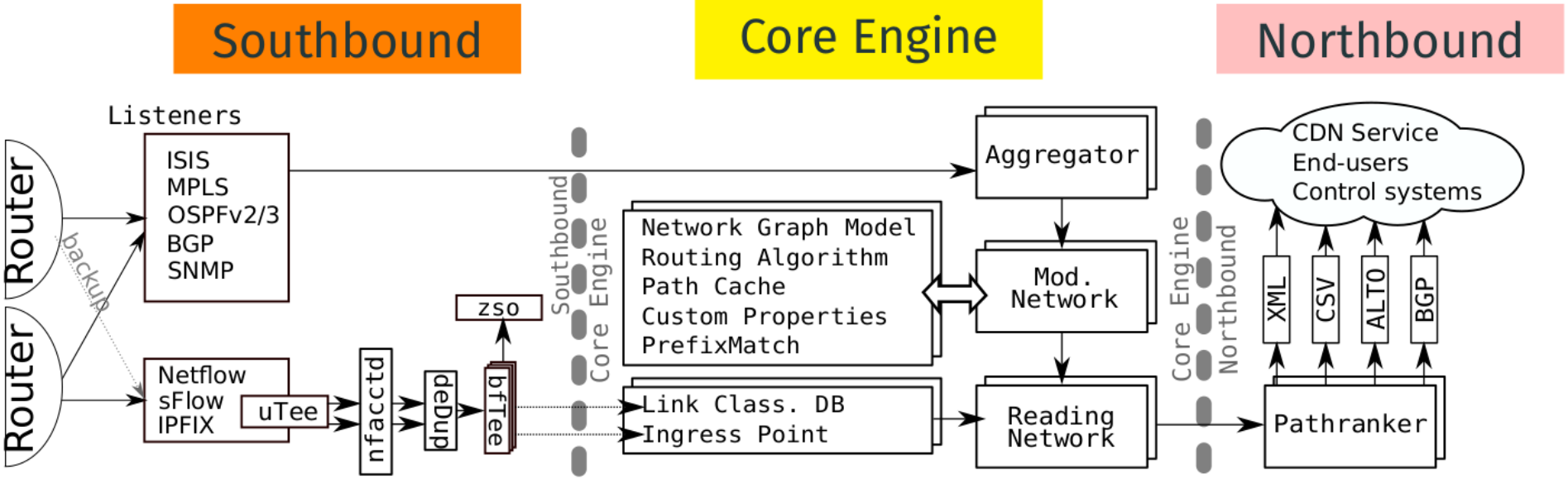
From a research idea to a production system



- Components design
  - RFC conforming input
  - Customizable output
  - Horizontally scalable
- Operational requirements

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

Architecture as of 2019



FlowDirector

Operational Experience

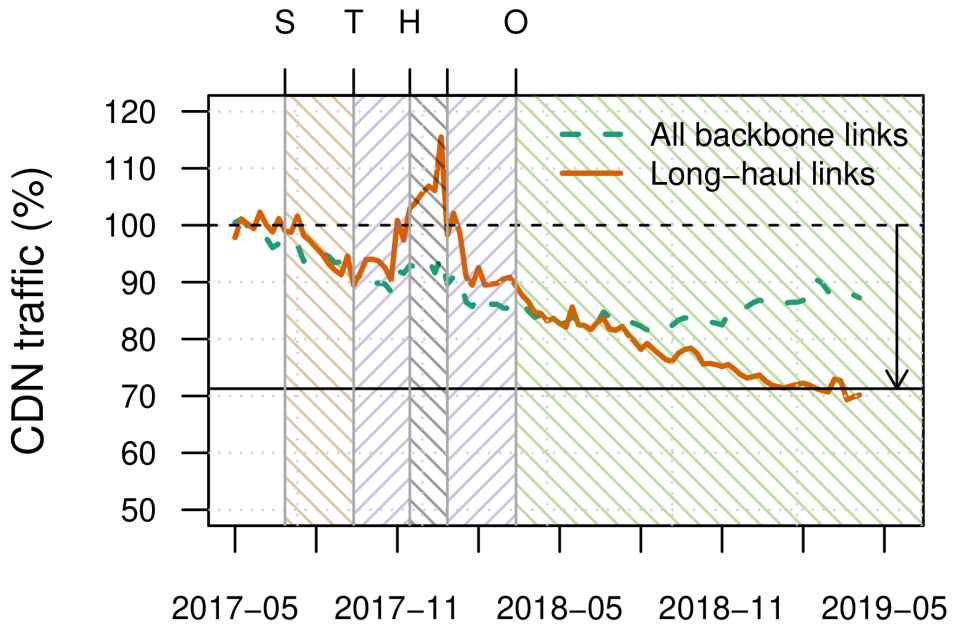
2-years experience with one cooperating hyper-giant

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

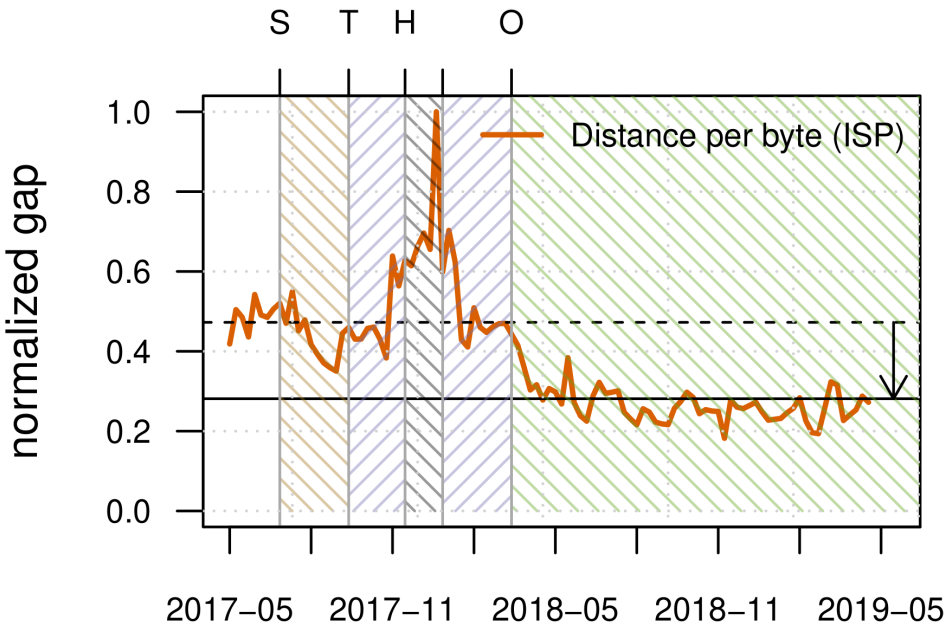
Combined with network planning:
30% reduction long-haul traffic



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

Benefits for the Hyper-giant

Distance as a proxy for latency:
40% reduction



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

ALTO experiences with ISP-CDN collaboration[4]

[4]Source: https://www.ietf.org/proceedings/96/slides/slides-96-alto-1.pdf)

Overview: ALTO implementation/deployment

1. Currently implemented
   - Base ALTO protocol (RFC7285) with all the provided features
     - IRD, NM,filtered-NM,CM, filtered-CM, ECS, EPS
     - Endpoints are not IP addresses but IP subnets (CIDR)
   - ALTO SSE
2. Deployed in production
   - NM and CM

Problem Statement

Scenario: CDN wants to deliver content to ISP customers

- Only paths from CDN caches towards customers of interest
- CDN caches embedded in foreign AS's
- How to group prefixes to form PIDs?

Network data accumulation

1. Routing protocols
   - IGP (IS-IS, OSPF, IBGP)
   - EGP (BGP)

   Links, routers, networks

2. Flow information
   - Netflow

   Ingress/Egress points

3. Network monitoring
   - SNMP

   Utilization, bandwidth, latency

Network Details

Large ISP Network to operate from:

- >900 Router
- >760k IPv4 Prefixes
  - >12k IGP Prefixes
  - >750k BGP Prefixes
    - >170k IBGP Prefixes
    - >580k EBGP Prefixes
- >20k IPv4 Ingress Prefixes
  - >950 Ingress Points
  - ~30% of public IPv4 Address Space

We have a live ALTO server providing guidance for CDN

Network Map

- Three different types of PID
  - Internal
  - External
  - OnNet
- External and OnNet prefixes are provided by Ingress Point Detection (IPD)
  - IPD: Tool that detects ingress points of external prefixes

Cost Maps

Three different Cost Maps:

- Hop Distance
- Path Weight
- Custom
  - No third party traffic on peering link
  - Only for specified ASN (OnNet ASNs)

Cost Calculation:

- Costs calculated between PIDs
- OnNet PIDs handled like Internal
- Outbound traffic is not considered
  - → No Egress Paths
  - → No Transit Paths

Statistics

- Updates every 5min
- Network Map
  - >250k Prefixes
  - >1700 PIDs
    - >750 Internal PIDs
    - >950 External PIDs
    - ∼ 15 OnNet PIDs
  - Average Map Sizes:
    - Map: ∼ 6 MB (∼ 1 MB compressed)
    - SSE Patch: ∼ 1.7 MB (∼ 282 KB compressed)
- Cost Map (Custom)
  - >1.3M PID pairs
  - Average Sizes
    - Map: ∼ 47 MB (∼ 5.6 MB compressed)
    - SSE Patch: ∼ 37.5 MB (∼ 5 MB compressed)

Problems

Long running map calculation process

- Cost Map calculation starts after Network Map is finished
- Network Map newer than current available Cost Map(s)
- Data inconsistency between Network and Cost Map possible

Limitation to IP addresses in ECS request

- RFC7285 states that input data for source/destination must be addresses (/32 for IPv4 and /128 for IPv6)
- Difficult requesting ECS for regions

Modification/Features

- Mechanism that publishes all maps together when last is ready
- Prefix support in ECS
- Timestamps and TTLs as meta fields in ALTO responses

Conclusion and future work

- FlowDirector
  - Opportunity to operate networks more efficiently
  - We enabled the first automated hypergiant-ISP collaboration
- ALTO: Next steps
  - Implementation/deployment experience documentation
  - No highest priority in terms of ALTO implementation, however, fully open to use our infrastructure as an evaluation/testing environment