Chair of Network Architectures and Services
Department of Computer Engineering
Technical University of Munich

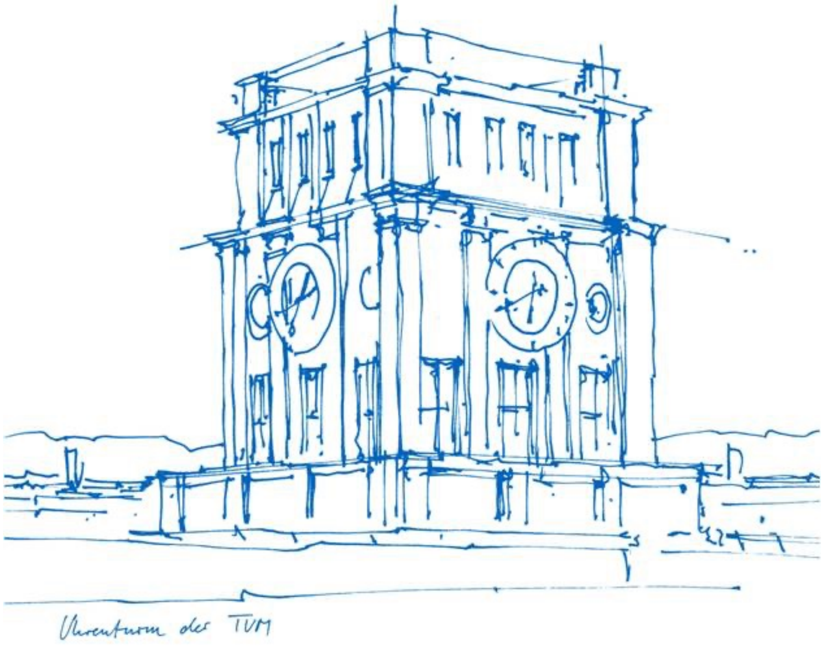
Research Infrastructures for Overlay Service Investigation

Georg Carle

carle@net.in.tum.de

http://www.net.in.tum.de/~carle

Acknowledgements:
All members of the Chair of
Network Architectures and Services



Outline /todo

Background

Research Infrastructures

- Internet Measurements
- Reproducible Experiments
- Large-scale Infrastructure for Computing/Communication Experimental Studies

Overlay Services

- MASQUE-Based Overlays
- Apple iCloud Private Relay
- Oblivious HTTP

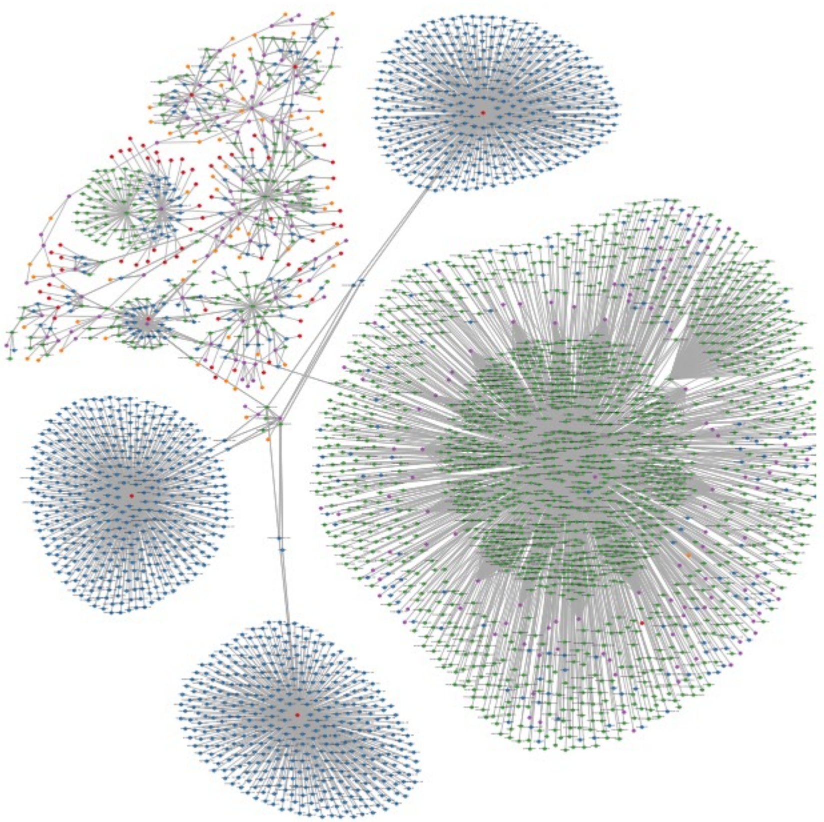
Conclusions

Chair of Network Architectures and Services
Department of Computer Engineering
Technical University of Munich

Internet-wide Measurements

Need to Understand the Internet

- The Internet consists of many network elements (e.g., routers) and server deployments (e.g. Web servers)
  - Different software, protocol stacks, configurations
- Challenge: characterisation of servers (e.g., trustworthy/malicius)
- Approach:
  - Large-scale measurements
  - Characterize behavior
  - Data:
    - L3/L4 data plane
    - Internet routing information
    - DNS
    - HTTP header values
    - TLS properties
    - X.509 certificates



TUM Infrastructure for Scanning the Internet

Local measurement infrastructure

- Scan servers (>10)
- Storage servers (currently 2)
- Analysis servers (currently 3)

Own autonomous system (AS56357)

- Dedicated border routers
- collect vantage point specific BGP information

Distributed measurement infrastructure

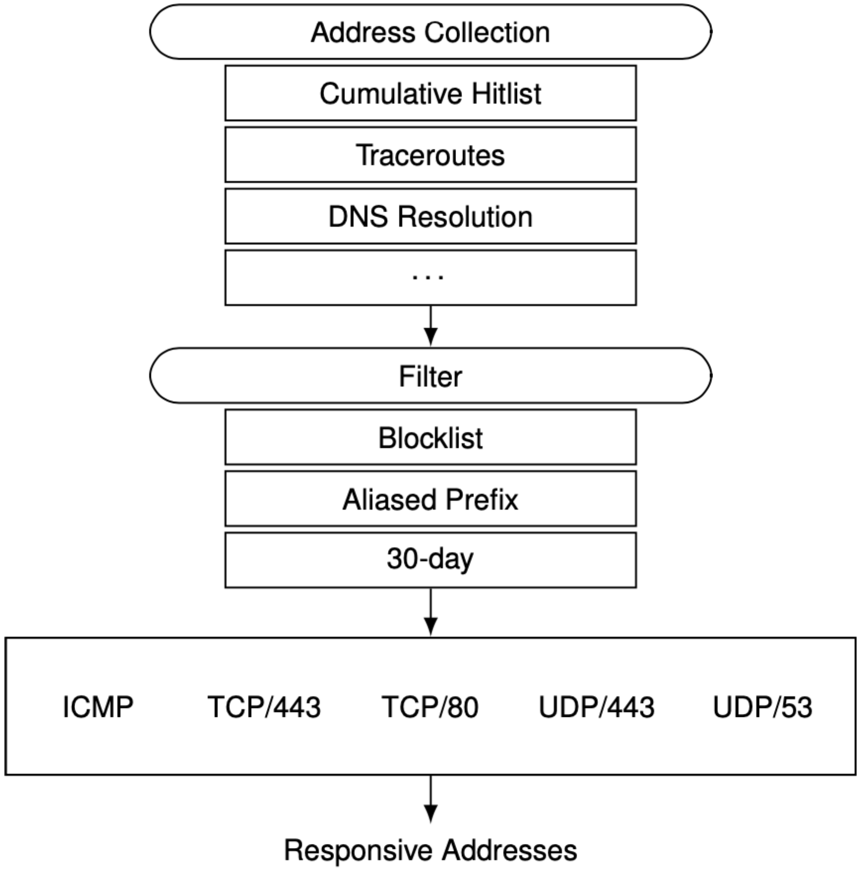
- Distributed root-servers and VMs within cloud providers
- RIPE Atlas anchor and measurements

Measurement targets

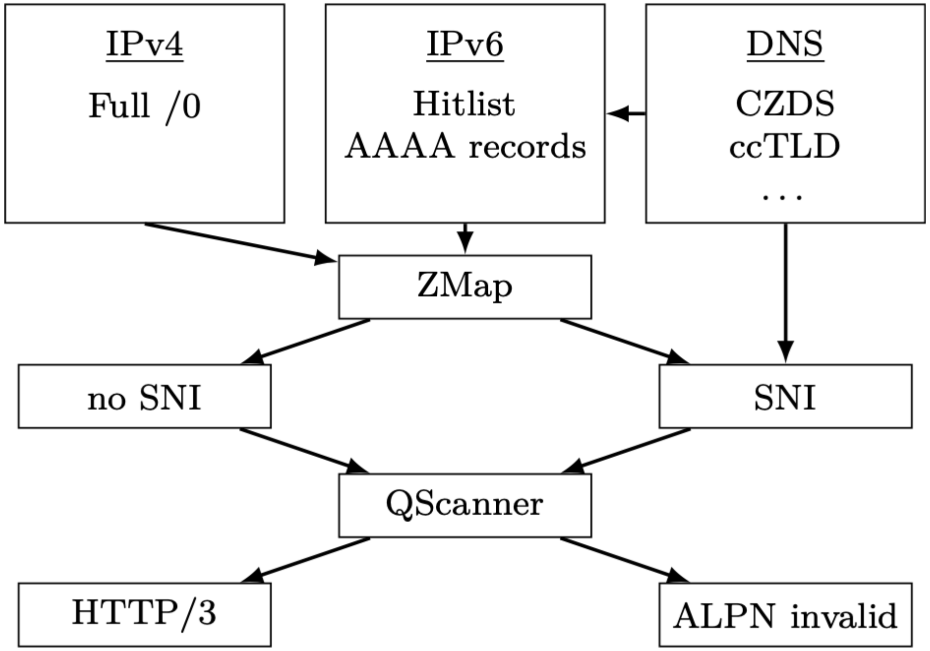
- DNS
- HTTP header values
- TLS properties
- X.509 certificates
- Routing information
- L3/L4 data plane

Scanning Global IPv6 and QUIC Deployments

NET.CIT.TUM IPv6 Hitlist



Global QUIC Scans



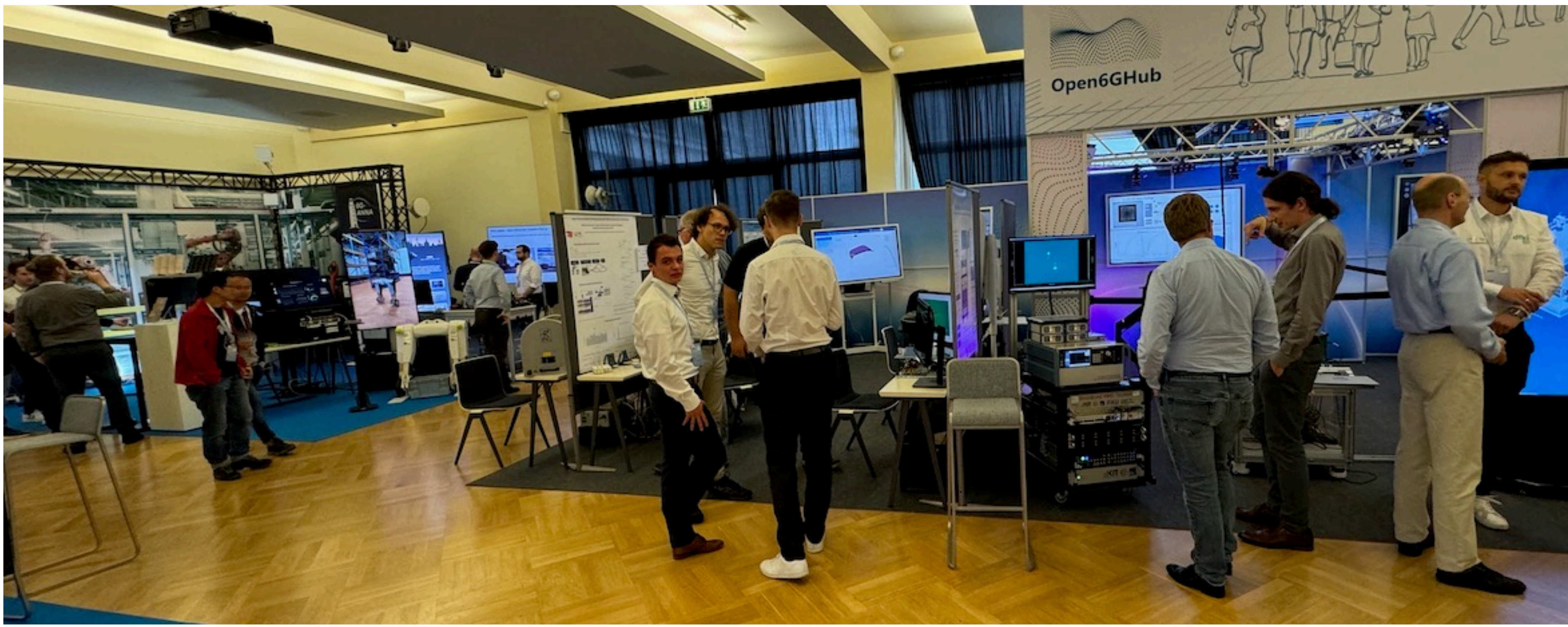
Lion Steger, Liming Kuang, Johannes Zirngibl, Georg Carle, Oliver Gasser, "Target Acquired? Evaluating Target Generation Algorithms for IPv6," in Proceedings of the Network Traffic Measurement and Analysis Conference (TMA), Jun. 2023. **Best Paper Award**

Chair of Network Architectures and Services
Department of Computer Engineering
Technical University of Munich

Reproducible Testbed Experiments

Large heterogeneity in Research Infrastructures

- Complexity of system architectures
- Purpose-built setups
- Difficulty to extend experiments of others

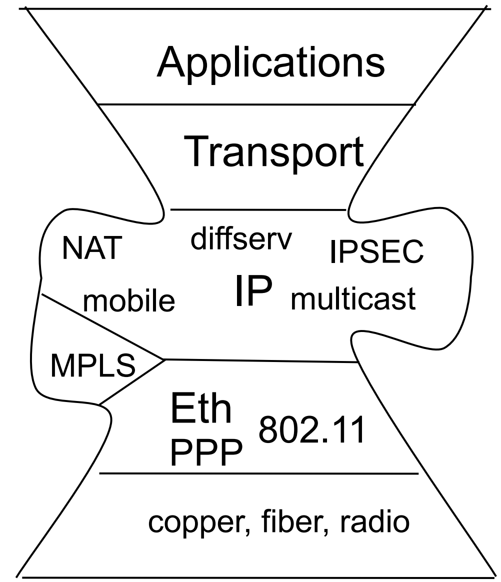


6G Conference, July 2024, Berlin

Challenge: Complexity

Protocol Stack

- TLS, QUIC, MASQUE
- UDP, TCP
- BGP, OSPF, VRRP, PIM
- IPsec, IKE, EAP
- IPv4, IPv6, Segment Routing
- VLAN, GTP, IP in IP, GRE, MPLS

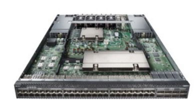


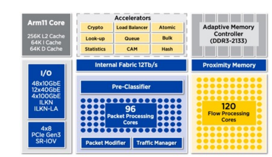
Complexity of Protocol Stack

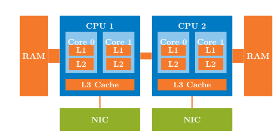
Complexity by Programmability

Complexity by Processing Architecture

Complexity by Software Architecture







Need: Reproducible Experiments

Viewpoints on Reproducible Research

ACM SIGCOMM MoMeTools - Workshop on Models, Methods and Tools for Reproducible Network Research
Georg Carle, Hartmut Ritter, Klaus Wehrle,
Karlsruhe, Germany, August 2003

ACM SIGCOMM Reproducibility Workshop
Olivier Bonaventure, Luigi Iannone, Damien Saucez
Los Angeles, USA, August 2017
[Rep17] Q. Scheitle, M. Wählisch, O. Gasser, T. Schmidt, G. Carle,
Towards an ecosystem for reproducible research in computer networking
Proceedings of the ACM SIGCOMM Reproducibility Workshop, 2017

Dagstuhl seminar 18412 "Encouraging Reproducibility in Scientific Research of the Internet", October 2018

- Despite 20 years since first workshop have passed, hard problems remain
- Current approaches
  - Artifact evaluation committees
  - Reproducibility badges
- Remaining problems
  - High effort for researchers to make research reproducible
  - High effort for members of artifact evaluation committee to validate reproducibility
  - Low robustness of experimental results due to insufficient documentation

Challenge: Inefficiency of Legacy Testbeds

- Testbeds are frequently created in the context of a specific project
- Purpose-built, highly complex systems
- After the project ends: two options

  (1) *conserve* the existing testbed

  - However, software must be kept up-to-date
  - Operational knowledge needs to be transferred to new scientists
  - Testbed nevertheless may be unfit to answer current questions

  (2) *repurpose* components to create testbeds for future projects

  - Ability to reproduce experiments usually lost
- Both options suboptimal concerning efficiency and sustainability

Approach

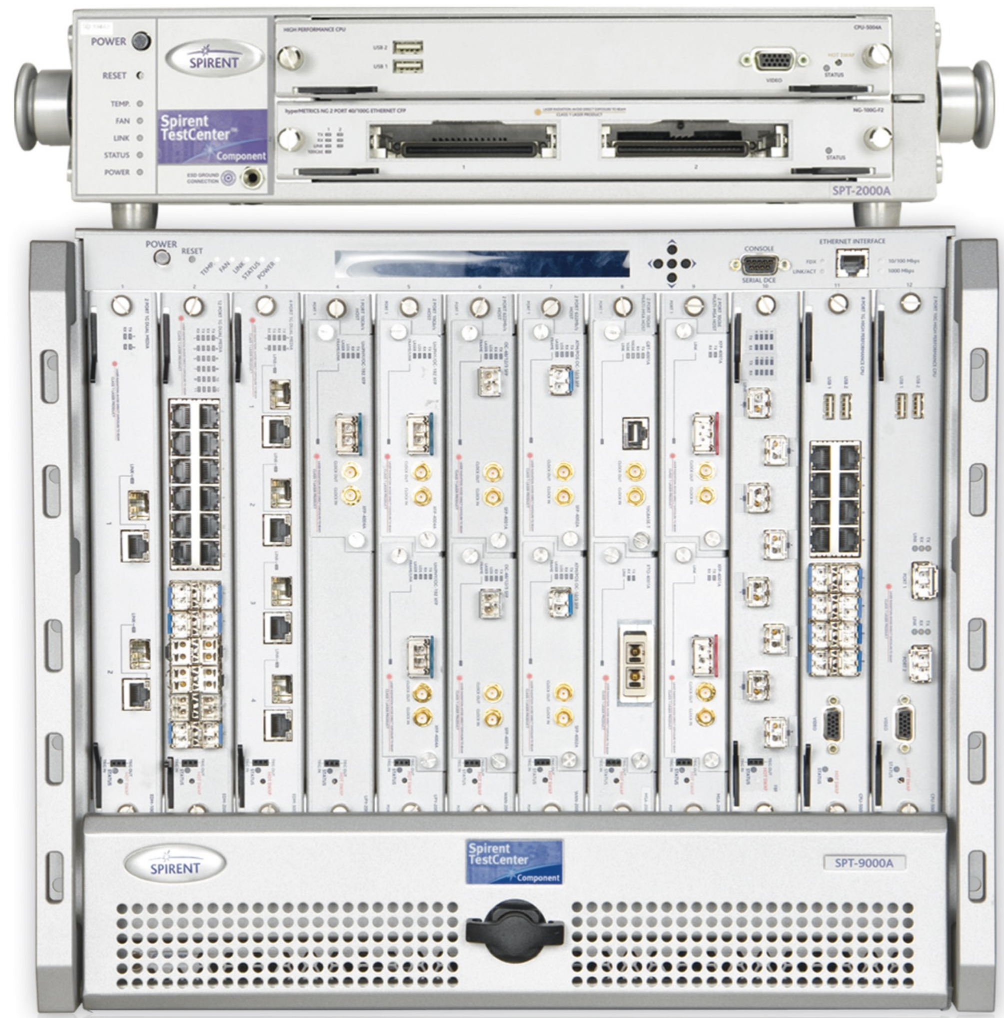
- Research Infrastructure for Reproducibility by Design

Hardware Traffic Generators

- Fast
- Precise

but

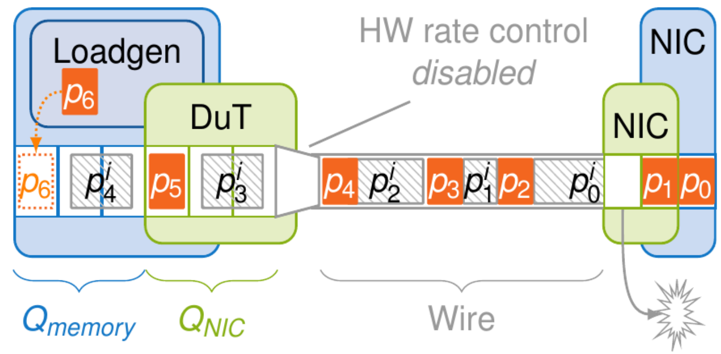
- Expensive
- Difficult to deploy
- Inflexible



Spirent traffic generator

MoonGen

- Inexpensive: Commercial Off-The-Shelf hardware
- Fast: DPDK for packet I/O, multi-core support
- Easy to deploy: simple software setup
- Flexible: user-controlled Lua scripts
- Precise
  - Timestamping: Utilize hardware features of commodity NICs
  - Rate control: Hardware features and software approach
  - Inter-packet spacing: gaps filled with invalid frames



[IMC15] Paul Emmerich, Sebastian Gallenmüller, Daniel Raumer, Florian Wohlfart, Georg Carle: MoonGen: A Scriptable High-Speed Packet Generator, ACM SIGCOMM Internet Measurement Conference (IMC), Oct. 2015

[ANRP17] Internet Research Task Force (IRTF) Applied Networking Research Prize, IETF-100, Nov. 2017, https://irtf.org/anrp

[ANCS17] Paul Emmerich, Sebastian Gallenmüller, Gianni Antichi, Andrew Moore, Georg Carle: Mind the Gap – A Comparison of Software Packet Generators, ACM/IEEE Symposium on Architectures for Networking and Communications Systems 2017

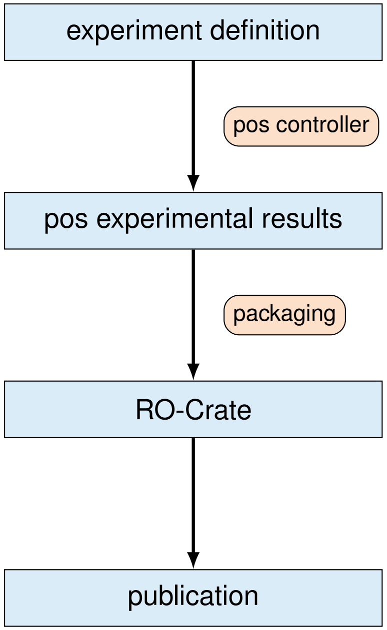
TUM Testbed for Reproducible Experiments

- Automated workflow using **pos** **p**lain **o**rchestrating **s**ervice [pos] workflow for reproducible experiments
- Throughput - packets per second, bytes per second, frame loss rate
- Latency - Median, average, worst case, percentiles, ...
- White-box - Hardware and software events; interrupts, cache misses

[pos] Sebastian Gallenmüller, Dominik Scholz, Henning Stubbe, Georg Carle, "The pos Framework: A Methodology and Toolchain for Reproducible Network Experiments," CoNEXT '21, Dec. 2021

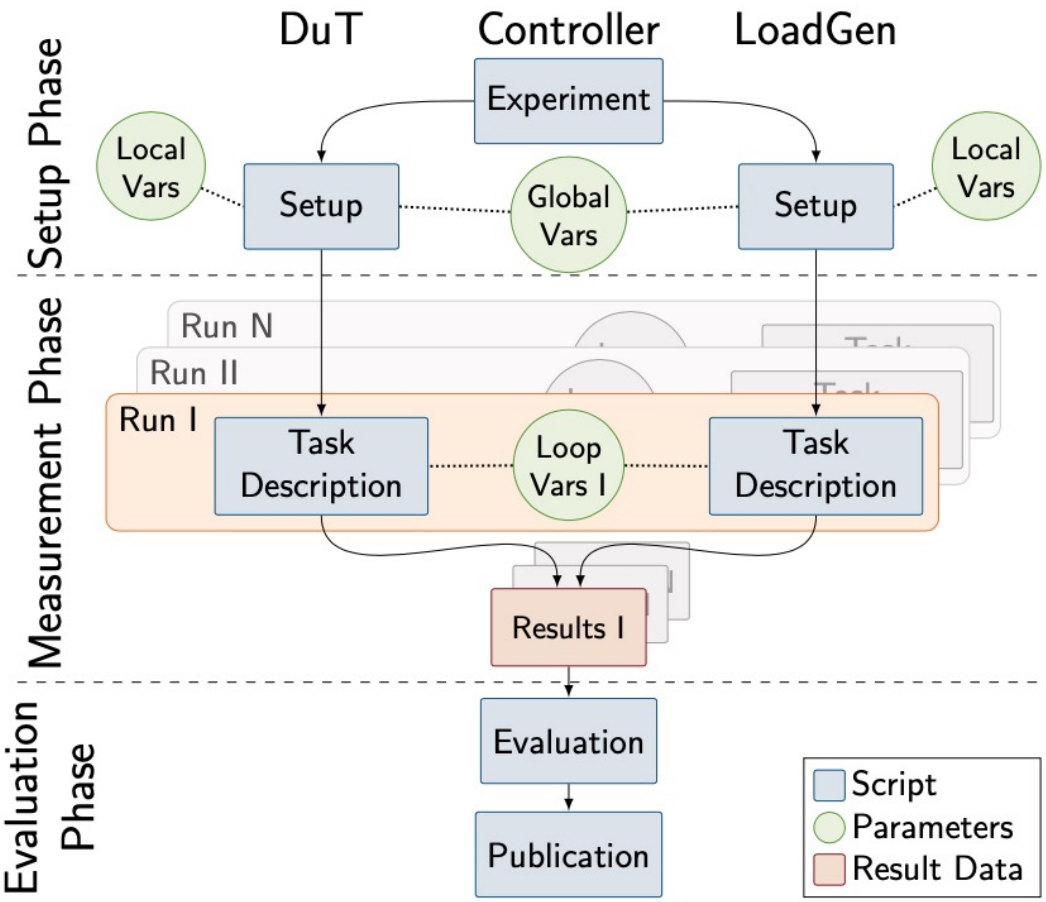
[RO-Crate] Eric Hauser, Sebastian Gallenmüller, Georg Carle, "RO-Crate for Testbeds: Automated Packaging of Experimental Results," in IFIP Networking Conference - SLICES Workshop, June 2024.

[SLICES] ESFRI - European Strategy Forum on Research Infrastructures; pos with TUM Baltikum Testbed: part of SLICES Research Infrastructure https://slices-ri.eu/



SLICES Benefits and Offering

Reproducibility by Design with pos workflow



pos workflow

Large-Scale Research Infrastructure

SLICES European Scientific Large-Scale Infrastructure for Computing/Communication Experimental Studies

ESFRI DIGIT Projects – Roadmap 2021

ESFRI: European Strategy Forum on Research Infrastructures

ESFRI Roadmap 2021 - Strategy Report on Research Infrastructures

https://www.esfri.eu/esfri-roadmap

https://roadmap2021.esfri.eu/media/1295/esfri-roadmap-2021.pdf

PAG 18

ESFRI PROJECTS

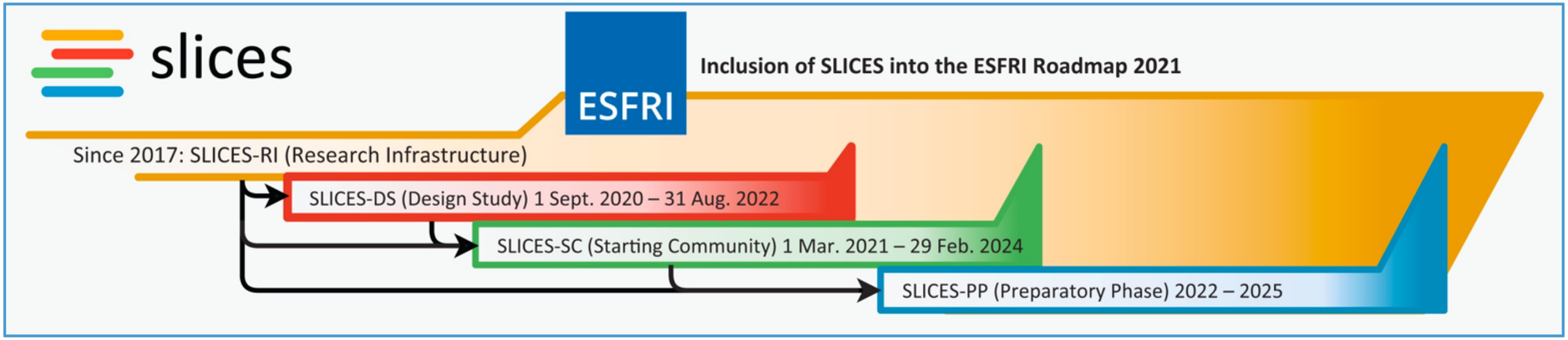
DIGIT

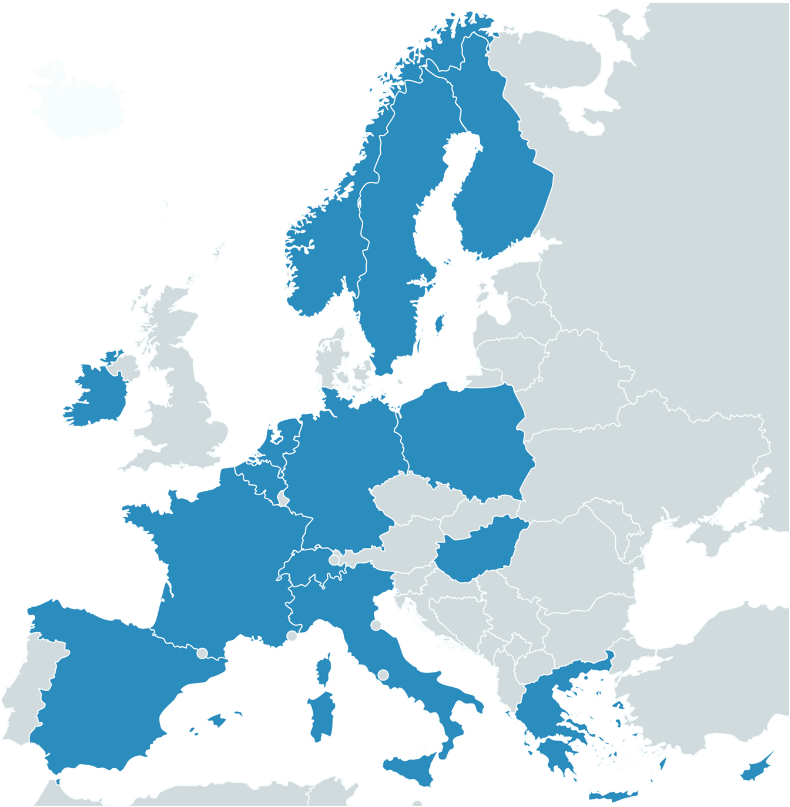
NAME	FULL NAME	TYPE	LEGAL STATUS (Y)	ROADMAP ENTRY (Y)	OPERATION START (Y)	INVESTMENT COST (M€)	OPERATION COST (M€/Y)
EBRAINS	European Brain ReseArch INfrastructureS	distributed	AISBL, 2019	2021	2026*	323.8	19.8
SLICES	Scientific Large-scale Infrastructure for Computing/ Communication Experimental Studies	distributed		2021	2024*	137.7	6.5
SoBigData++	European Integrated Infrastructure for Social Mining and Big Data Analytics	distributed		2021	2030*	130.5	5.0

SLICES is a **distributed Research Infrastructure** to design, develop and deploy the **Next Generation** of **Digital Infrastructures.**

SLICES-RI provides **specialized instruments** on research areas of Digital Infrastructures, by aggregating networking, computing and storage resources across countries, nodes and sites.

SLICES Partnership





SLICES Resarch Infrastructure Components and Services

Blueprints

- Historically: with blueprints, an unlimited numbers of accurate copies of plans can be produced
- SLICES blueprints: allows to reproduce software and hardware architectures at different sites
- First SLICES blueprint to deploy 5G cores and 5G RANs using OpenAirInterface and more, http://doc.slices-sc.eu/blueprint/
- More blueprints for Cloud, IoT, ... are being prepared

Reproducibility toolchain, including experiment orchestration with pos

Experiment portability with pos

Data management components

Educational material

Large-Scale Research Infrastructures

Testbed Research Infrastructures

- Can be attractive for networked systems experimental research
- Can provide large number of scientists access to specific resources
- Can provide tools that support reproducibility and portability
  - Experiment orchestration with pos
    - Reproducibility by design – guidance instead of experience
    - Portability of experiments – by supporting pos in different testbeds
- Data management components
  - FAIR: Findable, Accessible, Interoperable, Reusable
- Win-Win
  - Scientists: save time by not needing to build own research infrastructure, get access to resources, artifacts, results
  - Institutions: Large-Scale RI resource sharing more efficient and sustainable than research groups maintaining own testbeds
- Network effect: collaboration gets easier, which is beneficial for all

Chair of Network Architectures and Services
Department of Computer Engineering
Technical University of Munich

Secure Overlay Services

Secure Overlay Services

Apple Privacy-relevant technology
- iCloud Private Relay

Related standardisation
- IETF MASQUE
- 3GPP ATSSS

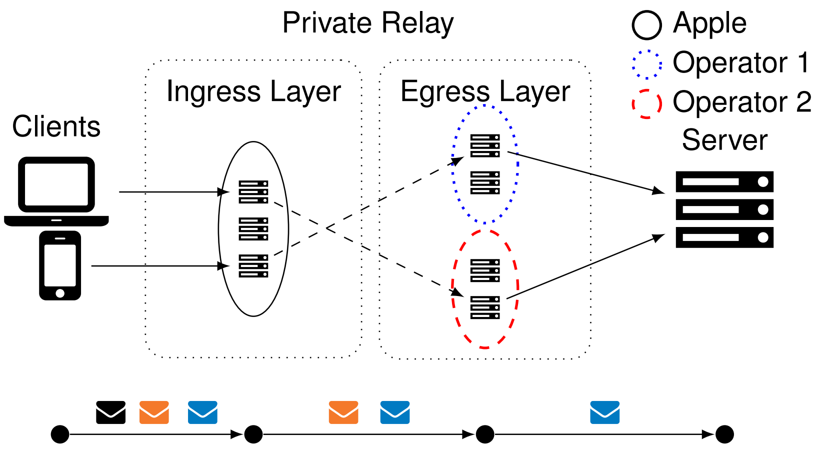
iCloud Private Relay Deployment analysis
- Egress proxies
- Ingress proxies

Cloudflare Privacy-relevant technology
- Oblivious HTTP

iCloud Private Relay

Main Goals

- "iCloud Private Relay is designed to protect your privacy by ensuring that when you browse the web in Safari, no single party – not even Apple – can see both who you are and what sites you're visiting."



What is it?

- Privacy Protection Service by Apple presented in June 2021 (WWDC'21)
- Available to all iCloud+ subscribers (cheapest plan is $0.99 per month)
- It uses IETF MASQUE to proxy the traffic

Publication

[IMC22] Patrick Sattler, Juliane Aulbach, Johannes Zirngibl, Georg Carle, "Towards a Tectonic Traffic Shift? Investigating Apple's New Relay Network," in Proceedings of the 2022 Internet Measurement Conference, Nice, Oct. 2022

5G Access Traffic Steering-Switching-Splitting (ATSSS)

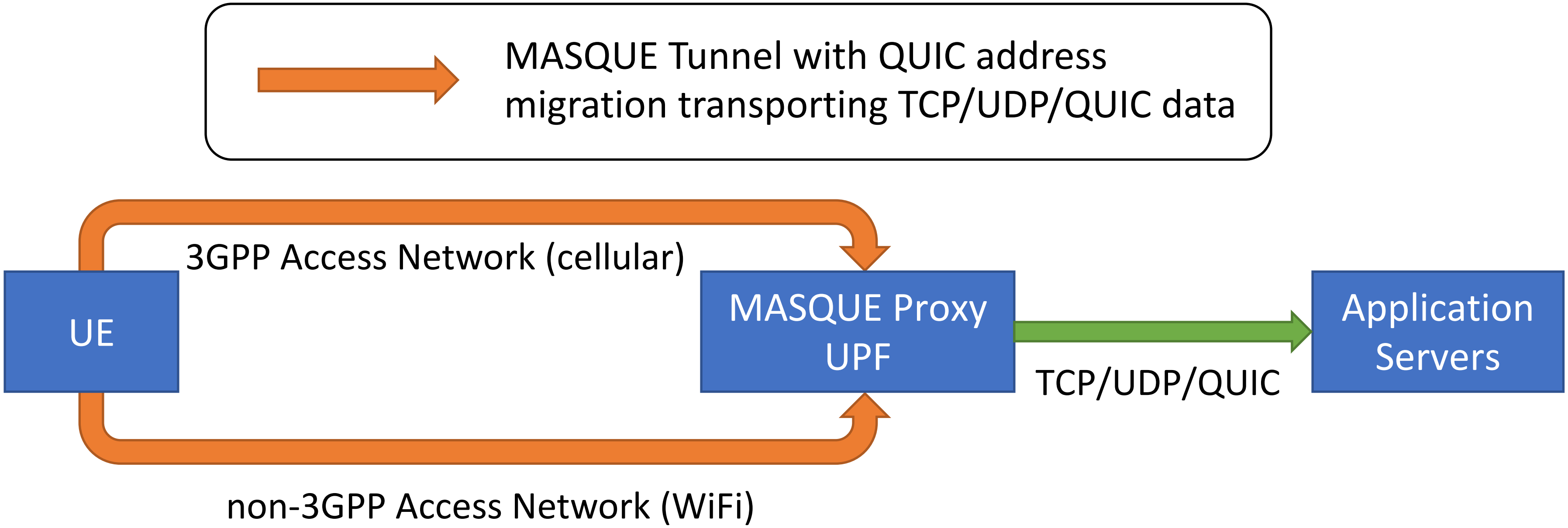
- Specification for 5G network access
- User Equipment (UE) might want to access a User Plane Function (UPF, a gateway) over a 3GPP network (e.g. cellular) and a non-3GPP network (e.g. WiFi) at the same time
- ATSSS provides the possibility to select/steer the path of specific data
- Connections to servers on the Internet can be established using Multipath-TCP (MPTCP) or Multipath-QUIC (MPQUIC)

[MPQUIC22] Yanmei Liu (Alibaba), Yunfei Ma (Alibaba), Quentin De Coninck (UCLouvain), Olivier Bonaventure (UCLouvain and Tessares), Christian Huitema (Private Octopus), Mirja Kühlewind (Ericsson), *Multipath Extension for QUIC,* Internet Engineering Task Force (IETF) Internet-Draft, Work in Progress, draft-ietf-quic-multipath-03, Oct. 2022

[MP21] Hongjia Wu, Simone Ferlin, Giuseppe Caso, Özgü Alay, Anna Brunstrom, *A Survey on Multipath Transport Protocols Towards 5G Access Traffic Steering, Switching and Splitting*. IEEE Access, 9 Dec 2021, 9:164417-39

Use Case: Multipath-Tunnelling in ATSSS

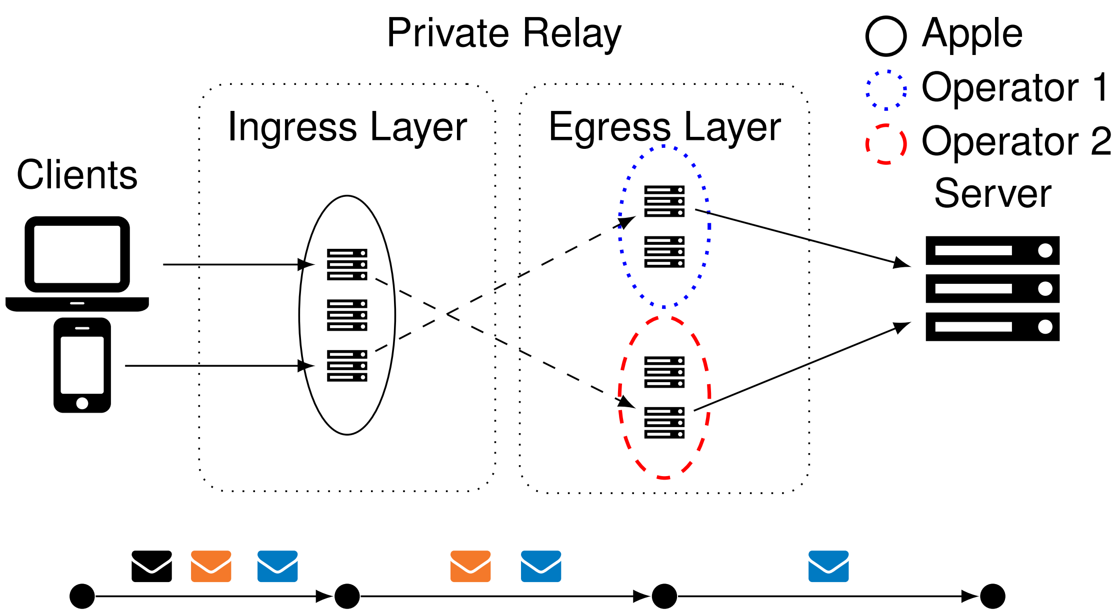
- QUIC provides connection migration between paths, ideal for traffic steering, MPQUIC enables simultaneous usage of paths
- MASQUE can tunnel both TCP and non-TCP data over one multi-path QUIC tunnel



[QUICWG20] M. Pirau, O. Bonaventure, Q. De Coninck, S. Dawkins, M. Kuehlewind, M. Amend, A. Kassler, Q. An, N. Keukeleire, S. Seo, 3GPP Access Traffic Steering Switching and Splitting (ATSSS) - Overview for IETF Participants, Internet-Draft, Work in Progress, draft-bonaventure-quic-atsss-overview-00, 2020

iCloud Private Relay

- Two separate relay node types
- Client-facing ingress proxies operated by Apple
- Server-facing egress proxies operated by third parties
- Clients connect via *mask.icloud.com* to ingress nodes are located around the world
- Private relay users and clients are verified and abuse is prevented according to Apple
- Only ingress proxy knows the client IP address,
  only egress proxy knows the server IP address



iCloud Private Relay

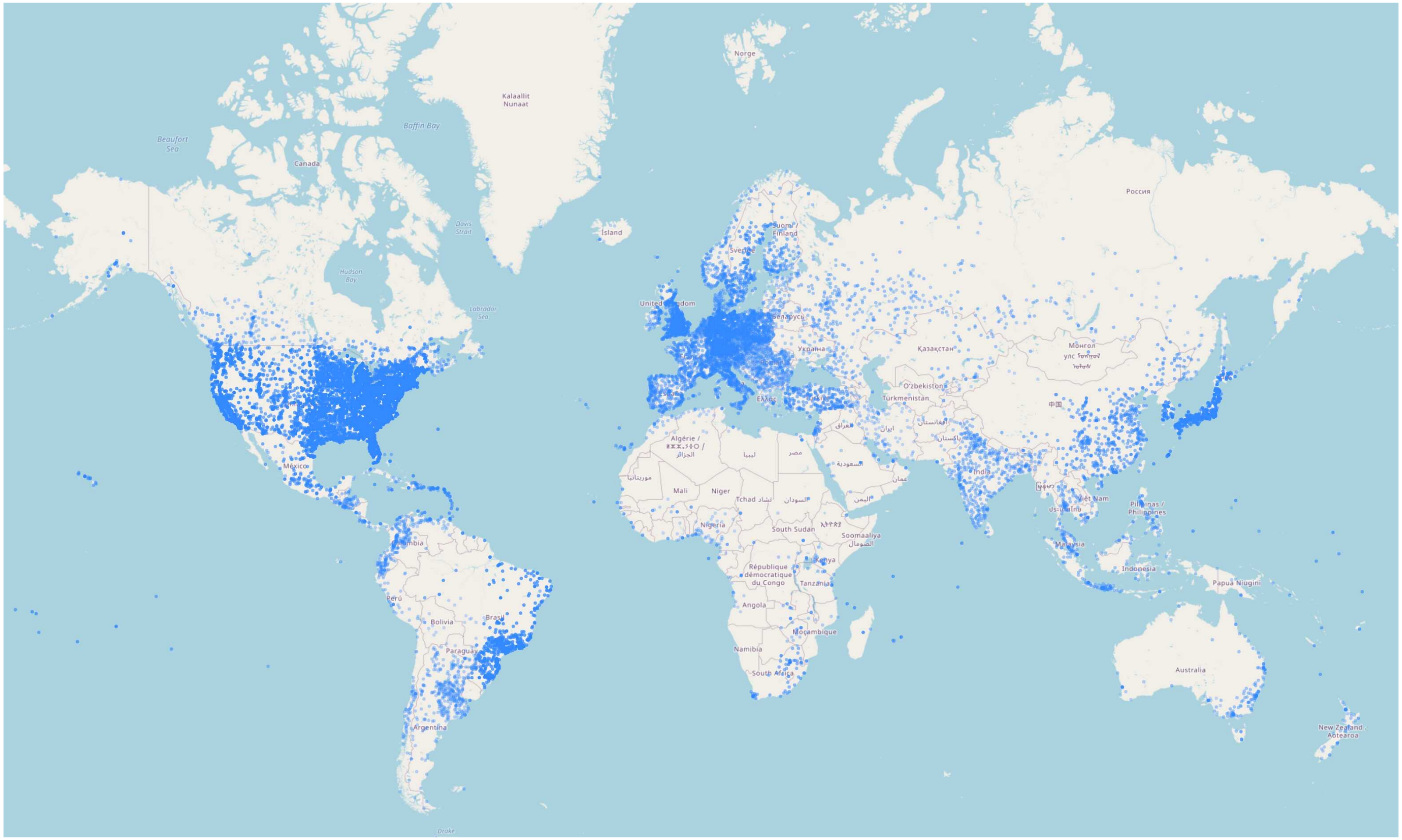
Current Deployment - Egress Proxy Operation

- Apple provides egress addresses together with their representing location
- Clients get egress addresses assigned representing their location
- Can be used to perform geo-blocking or location-based ad targeting
- Akamai provides the largest number of addresses
- No indicator for other services using $\text{Akamai}_{\text{PR}}$ AS
    - BGP announcements started in May 2021
- Current deployment focus is on developed countries

	Subnets	BGP Pfxs	IP Addr.	CCs
Akamai_{PR} (AS36183)	9890	301	57 589	236
Akamai_{Est} (AS20940)	1602	1	5100	24
Cloudflare (AS13335)	18 218	112	18 218	248
Fastly (AS54113)	8530	81	17 060	236

Geolocation of egress subnets

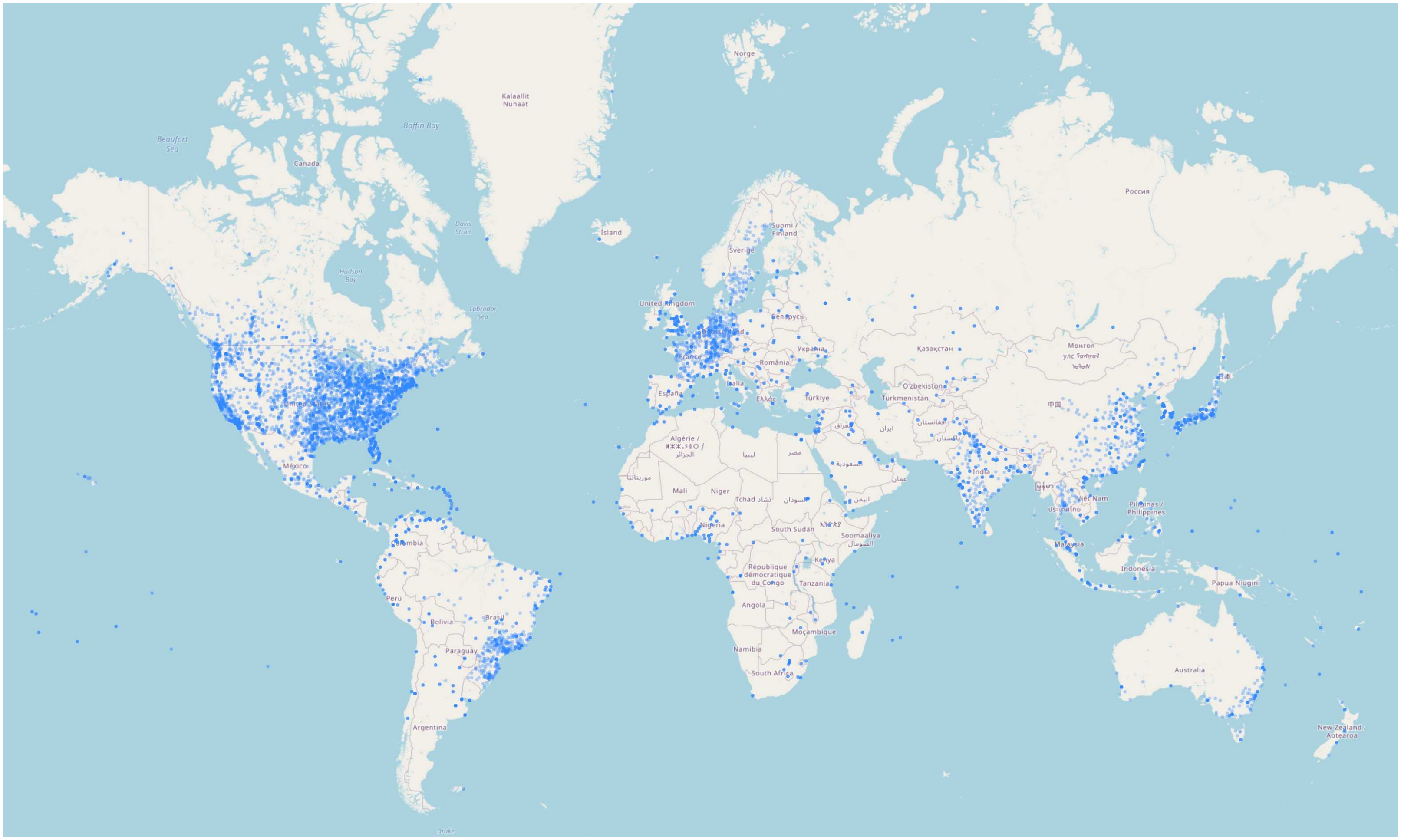
Akamai$_{PR}$ and Akamai$_{Est}$



iCloud Private Relay

Geolocation of egress subnets

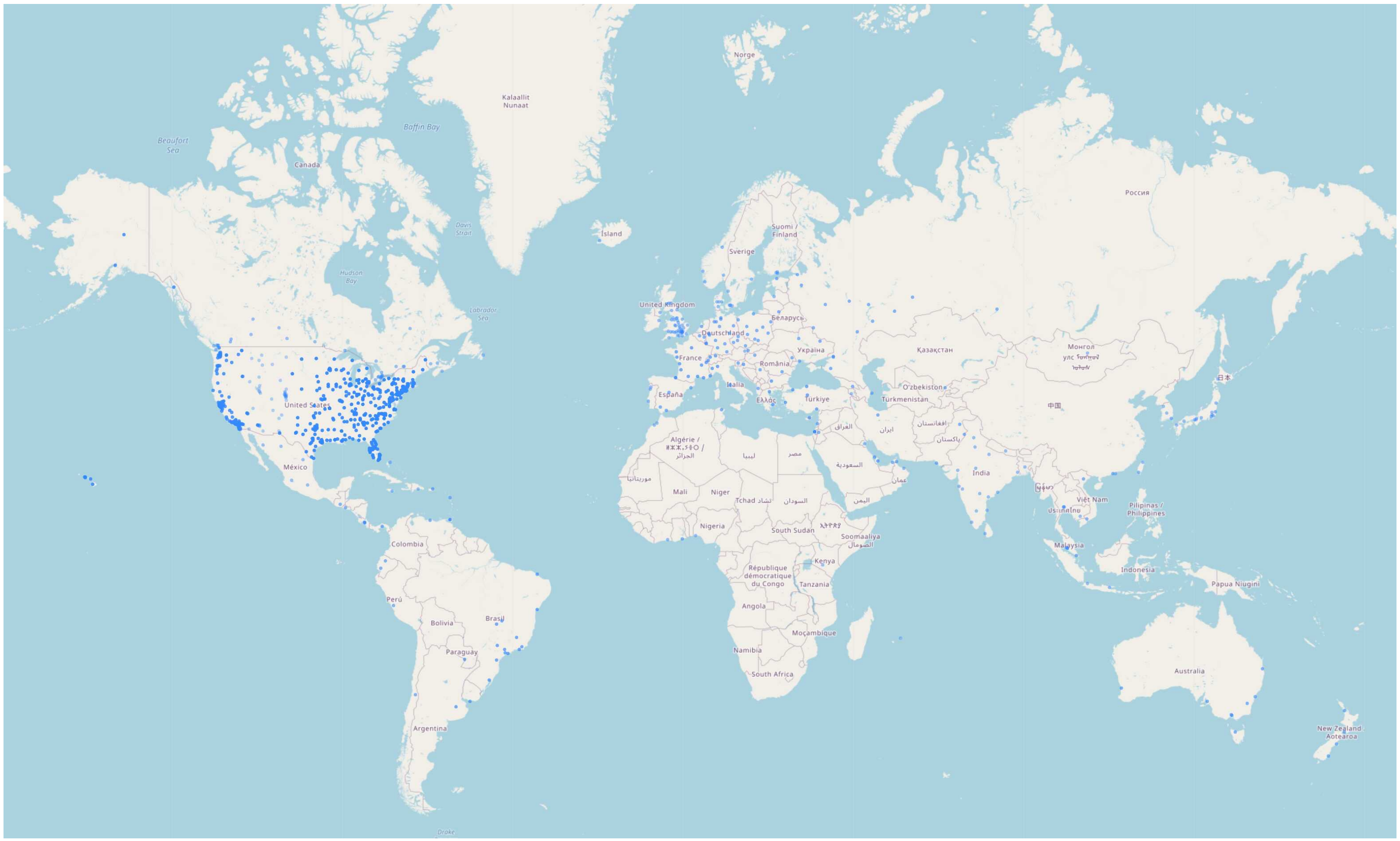
Cloudflare



iCloud Private Relay

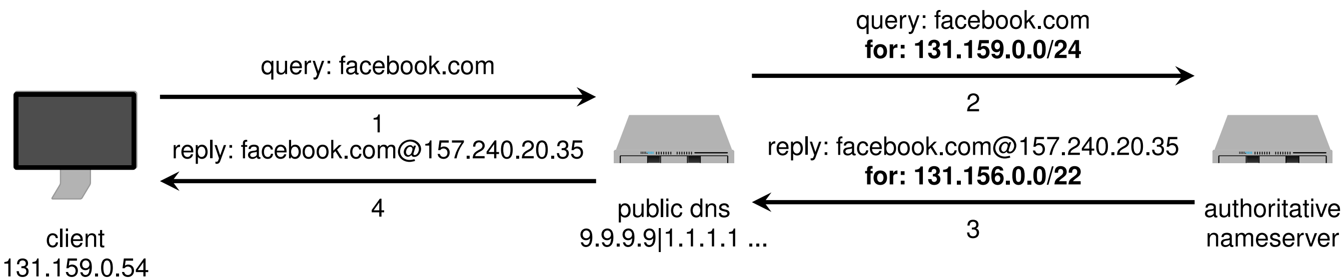
Geolocation of egress subnets

Fastly

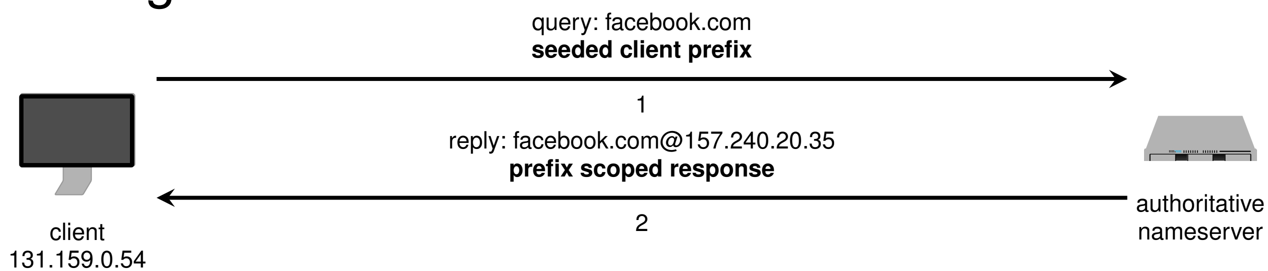


Current Deployment - Ingress Addresses

- Information on Ingress addresses is not publicly available
- If QUIC-based connections to ingress fail a fallback to TCP/TLS1.3 is used
- We performed client subnet enumeration scans using the DNS extension *EDNS Client Subnet (ECS),* RFC 7871, to obtain ingress IP addresses to connect to QUIC ingress proxies and TCP fallback proxies
- With ECS a subnet can be attached to a DNS query and be used by the name server for prefix-scoped answers
- ECS: resolver includes client subnet in DNS query



- ECS scanning: scanner sends ECS Parameters for different subnets directly



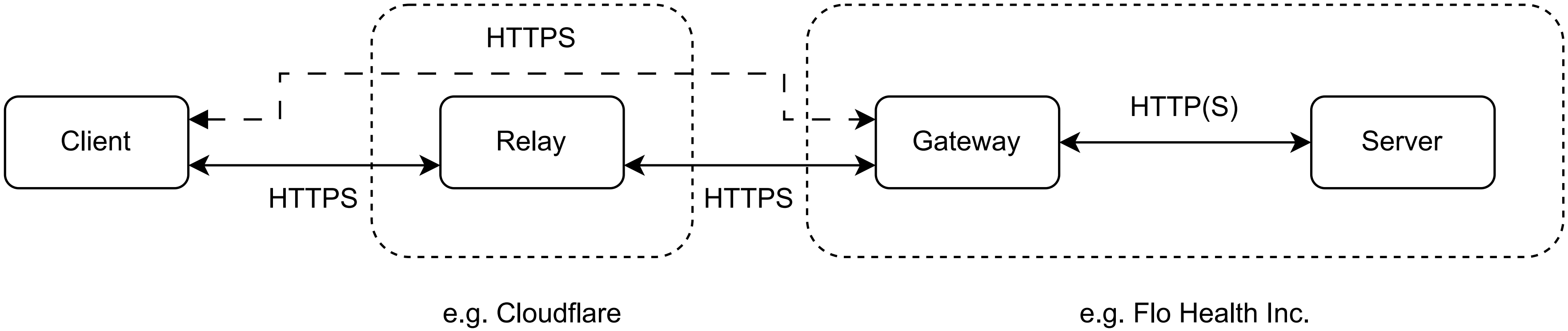
Current Deployment - Ingress Addresses

- Information on Ingress addresses is not publicly available
- Results of our ECS enumeration scans to obtain ingress IP addresses
- 44% increase of ingress addresses over ten months
- In addition to Apple's AS, we also find ingress relays in $\text{Akamai}_{\text{PR}}$
- More than 70% of ingress relays are inside $\text{Akamai}_{\text{PR}}$

	Total	Apple		Akamai_{PR}	
Jan	1188	365	30.6 %	823	69.4 %
Feb	1200	355	29.5 %	845	70.5 %
Mar	1292	347	26.9 %	945	73.1 %
Apr	1586	349	22.0 %	1237	78.0 %
			...		
Oct	1713	475	27.7 %	1238	72.3 %

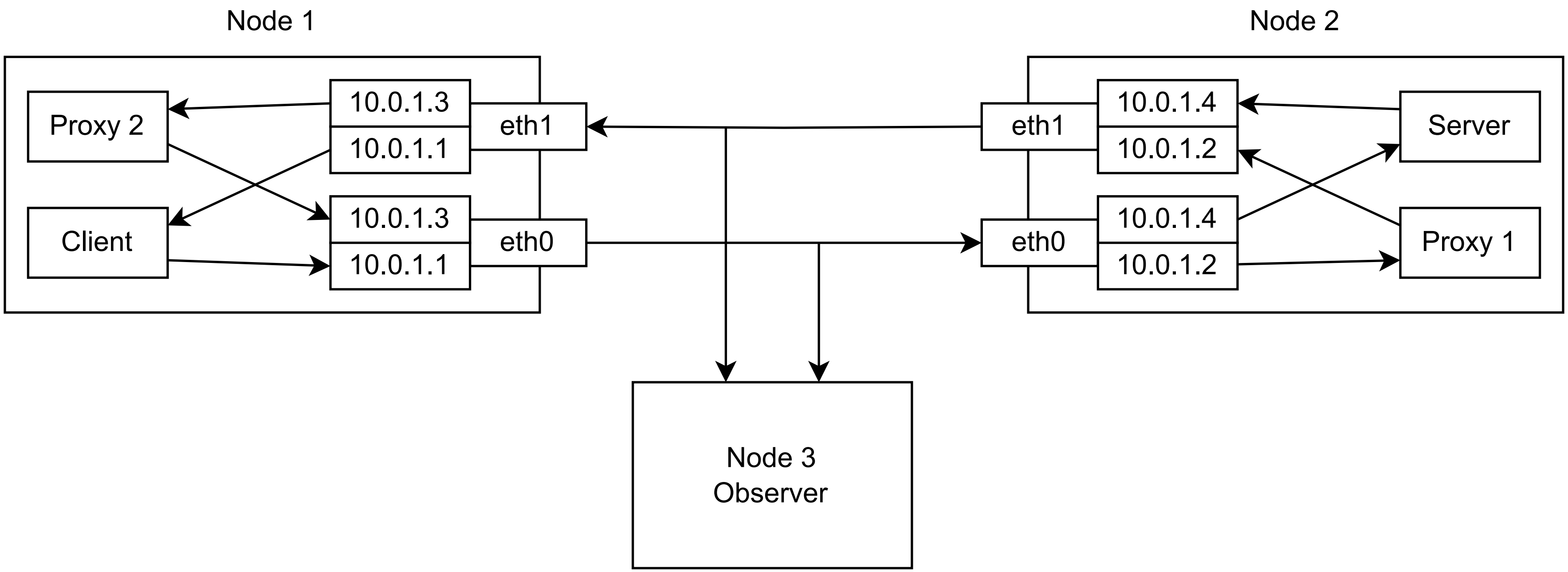
Oblivious HTTP

- Two-stage proxy architecture for HTTP, Standards Track RFC 9458
- Proxy 1 (Relay) operated by different organization than Proxy 2 (Gateway) and the destination server
- Proxy 1 (Relay) prevents server knowing client IP address
- Secure tunnel between client and Proxy 2 (Gateway) using Hybrid Public Key Encryption (HPKE) prevents Relay accessing payload
- No single organization can correlate the communication partners



Testbed Setup

- Investigating MASQUE and Oblivious HTTP
- Two Proxy Scenario:



Conclusions - Overlay Services

- Privacy relays have a large potential user base
- Traffic shifts affect multiple stakeholders
- Obtained ingress IP addresses may be used to evaluate and handle this type of traffic
- Akamai is the largest ingress and egress operator (using its dedicated $\text{Akamai}_{\text{PR}}$ AS)

Publication of our iCloud Private Relay measurement data:

- All scan results including detailed ECS responses
- Archive of egress IP addresses: https://relay-networks.github.io/

More Overlay Services can be expected

Conclusions - Large-Scale Research Infrastructures

Testbed Research Infrastructures

- Can be attractive for networked systems experimental research
- May provide large number of scientists access to specific resources
- Should provide tools that support reproducibility and portability
  - Experiment orchestration with pos
    - Reproducibility by design – guidance instead of experience
    - Portability of experiments – by supporting pos in different testbeds
- Data management components
  - FAIR: Findable, Accessible, Interoperable, Reusable
- Win-Win
  - Scientists: save time by not needing to build own research infrastructure, get access to resources, artifacts, results
  - Institutions: Large-Scale RI resource sharing more efficient and sustainable than research groups maintaining own testbeds
- Network effect: collaboration gets easier, which is beneficial for all

Questions?