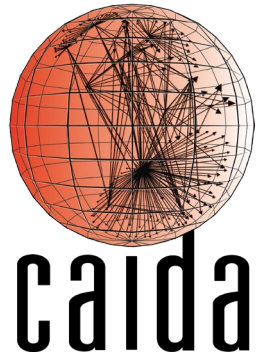
From Darknet to Dual-Stack: Challenges and Advances in Internet Telescope Infrastructure

Ricky Mok

CAIDA/UC San Diego

IIJ May 18, 2026



Network telescopes

- Leverage large blocks of *unused* IP address space (**darknet**) to collect *unsolicited* Internet traffic, namely Internet Background Radiation (IBR)
- IBR consists of
  - Internet-wide scanning traffic
    - “Legit” scanning (e.g., censys, academic research)
    - Malicious (e.g., malware, attackers seeking vulnerable hosts)
  - Backscatter generated by randomly-spoofed DoS attacks
  - Misconfigurations

Network telescope deployment

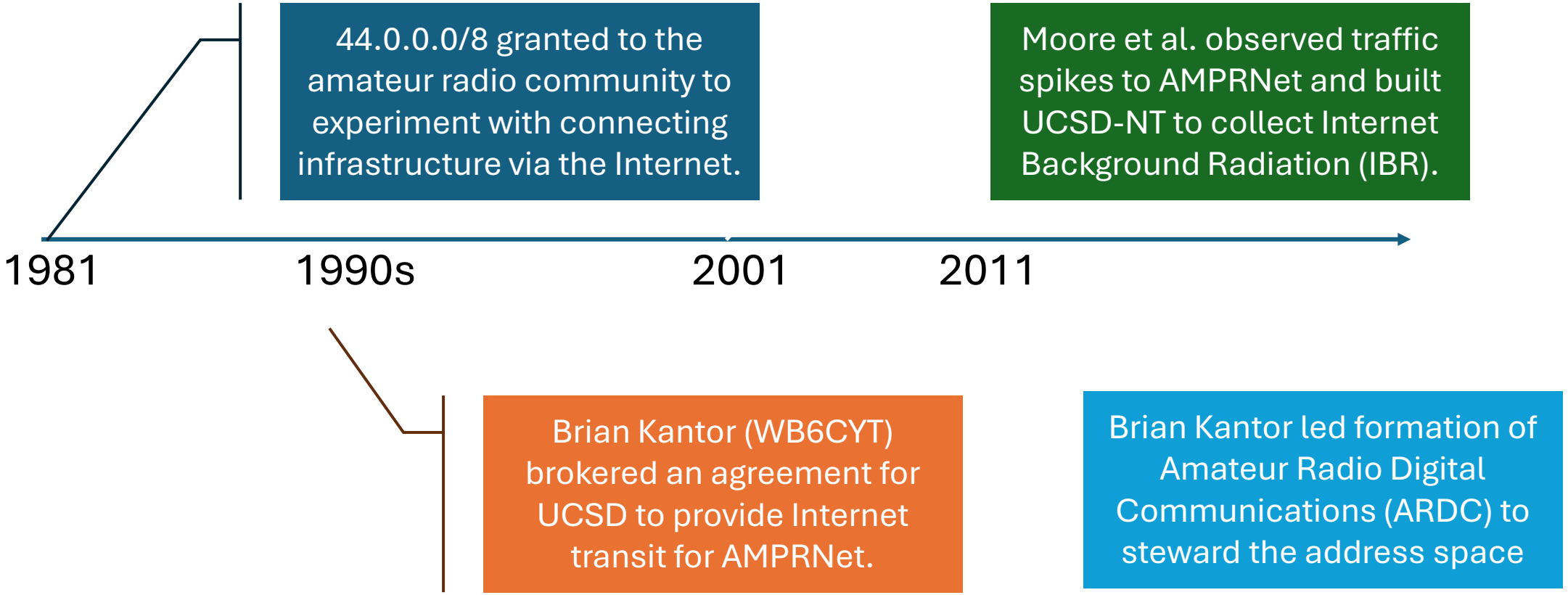
- Multiple (international) efforts in deploying network telescopes for cybersecurity research

Network telescopes	Address spaces	Countries
UCSD-NT	/9+/10	US
Merit ORION	~7x /16	US
NICTER	/17, /18/, 2x/20	JP
SURF	/16	NL
Politecnico di Torino	3x /24	IT

UCSD-NT

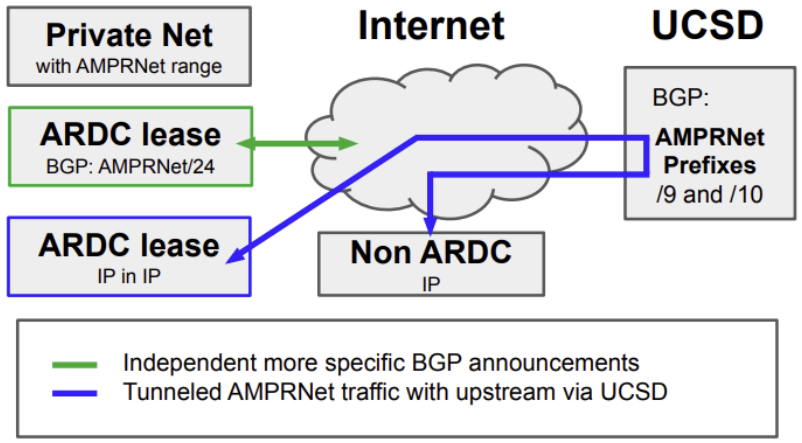
- The world's largest network telescope
  - /8 → /9+/10 (2019)
    - To date, ~10.3M dark IPv4 addresses
- 100-150 GB of gzip-compressed pcap per hour
  - ~3-4 Billions packets
- CAIDA hosts the last 30 days of pcap and historical flow aggregated data

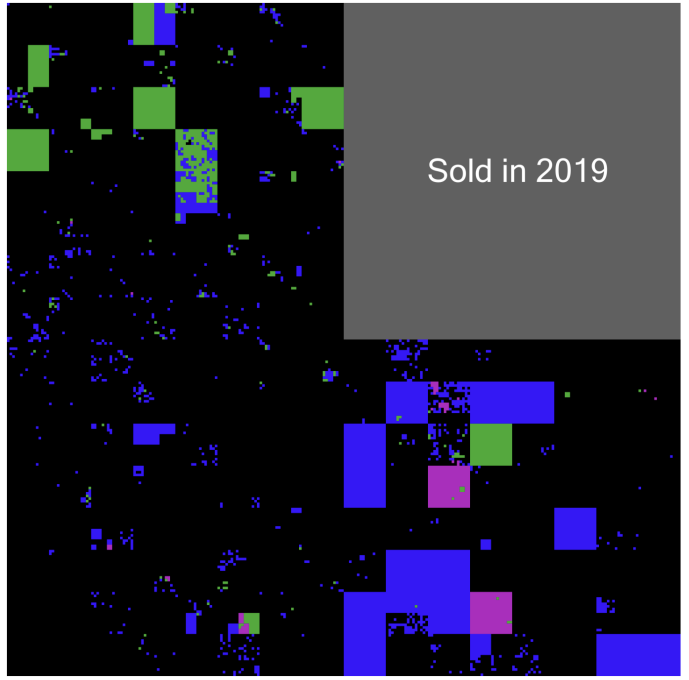
A brief history of UCSD-NT



Unique characteristic of UCSD-NT

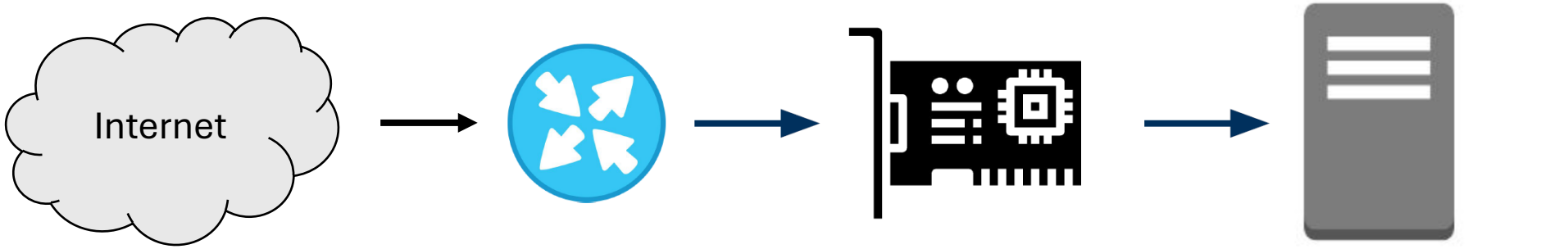
- Share address space with ARDC
  - ARDC allocates/leases address spaces to the amateur radio community
  - UCSD-NT queries ARDC's RESTful API to filter out legitimate traffic
- UCSD-NT leverages multicast to support capturing near-real-time traffic streams in a VM



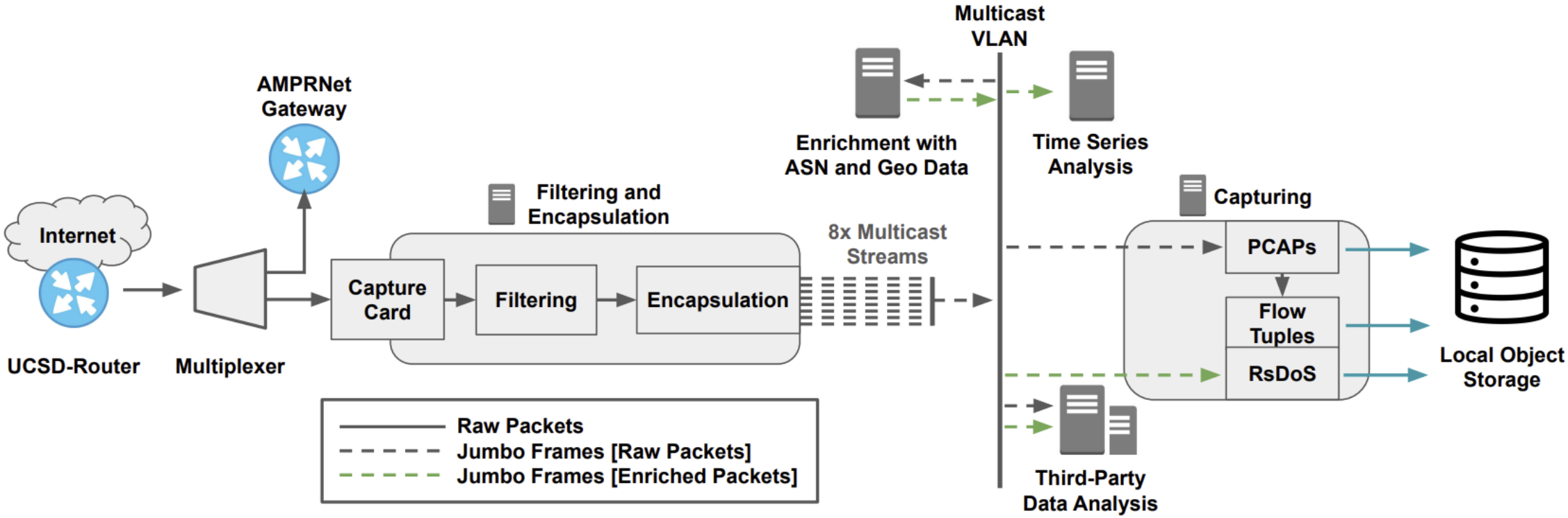


A. Männel et al. "Lessons Learned from Operating a Large Network Telescope", In ACM SIGCOMM 2025.

The architecture of UCSD-NT



The architecture of UCSD-NT

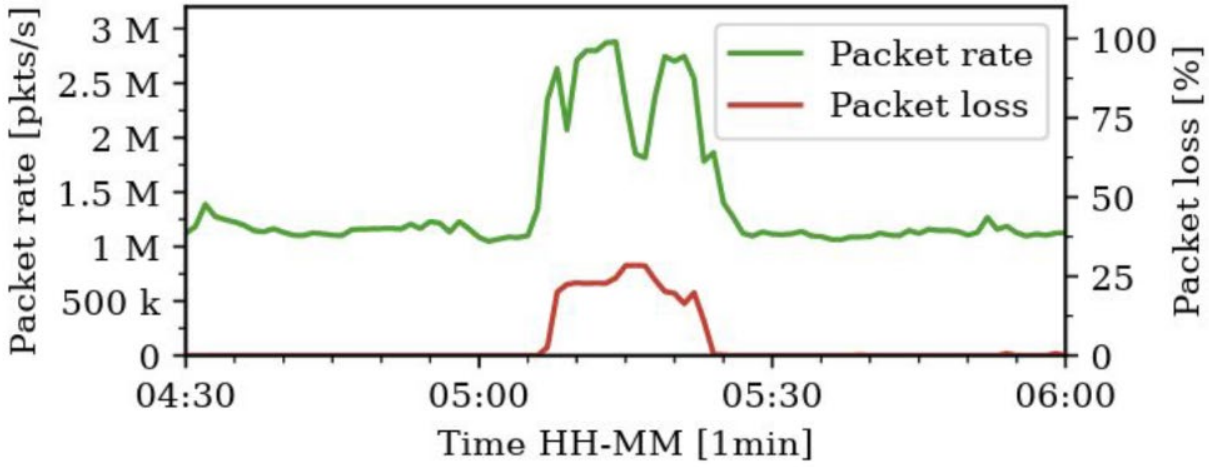


Complexity brings challenges

- Packet lost due to growing traffic volume
- Increased dynamics of AMPRNet
  - Communication and operator errors
- Internet routing
  - “Honest” BGP hijacks
  - Traffic engineering in some ISPs

→Missing traffic

→Accidentally capture legitimate traffic



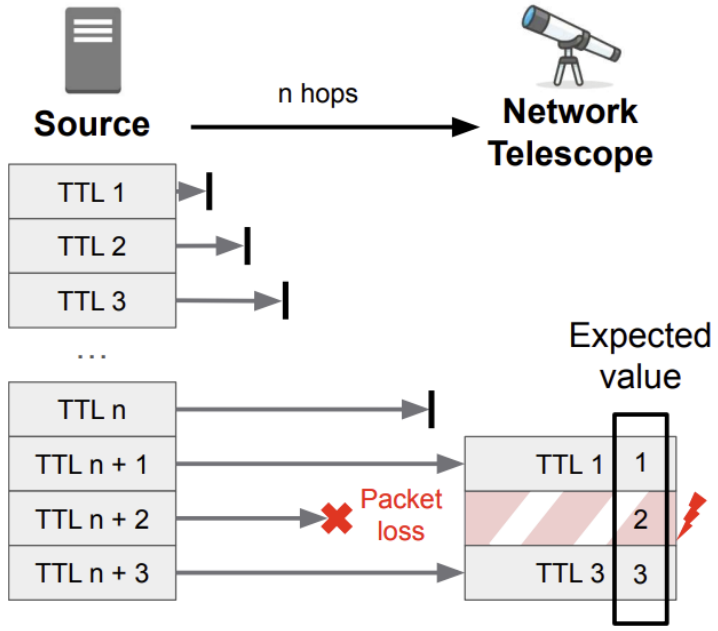
Improve monitoring capability

- Capture system telemetry
- Provide information to data users about the reliability of the data
- Lower our admin’s response time to failures

Demo

“Zero-cost” monitoring

- Leverage Ark and existing scanning campaigns to identify packet loss
  - Alphastrike
  - Leitwert
- Missing expected packets likely indicate packet loss in UCSD-NT
- New SmartNIC-based network monitoring is undergoing testing to reduce packet loss



AI-ready UCSD-NT

- Researchers investigate how to use ML/AI to extract insights from IBR
  - Detect traffic anomalies: Dark-Tracer [Han2022], DarkSIM [Gao2024]
  - Identify scanners with similar behavior: DarkVec [Gioacchini 21]
- Pcap data is difficult for ML/AI to process
  - Lack of “labels”
  - The community has no “reference data” for training
  - Hard to extract subset of data

  → Researchers invent different evaluation methods with their own datasets

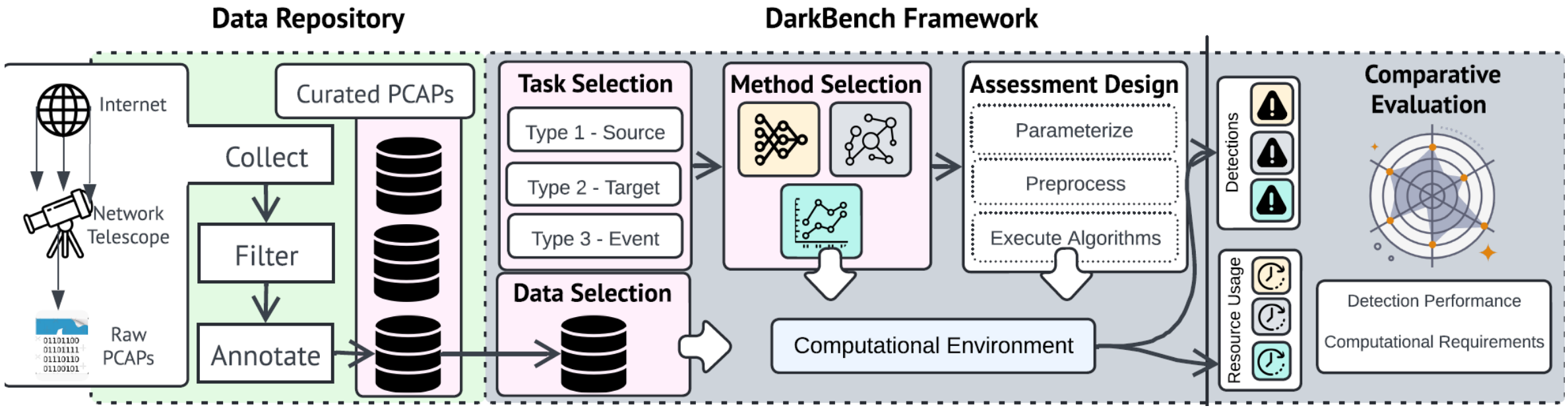
[Han2022] C. Han et al , ''Dark-TRACER: Early Detection Framework for Malware Activity Based on Anomalous Spatiotemporal Patterns,'' IEEE ACCESS, 2022.
[Gao2024] M. Gao et al., "DarkSim: A similarity-based time-series analytic framework for darknet traffic", In ACM IMC, 2024.
[Gioacchini 21] L. Gioacchini et al. “Darkvec: Automatic analysis of darknet traffic with word embeddings”, In ACM CoNEXT 2021.

CANIS

- NSF project: “Curated AI-ready Network telescope datasets for Internet Security” (OAC-2531134)
- We offer
  - Label traffic based on packet fingerprints, known scanners
  - “Reference datasets” contain known events (e.g., onset of Mirai)
  - Apache Parquet data format to facilitate “data science” researchers to use the data
  - Facilitate researchers to use “supercomputer” to analyze the data
  - DarkBench

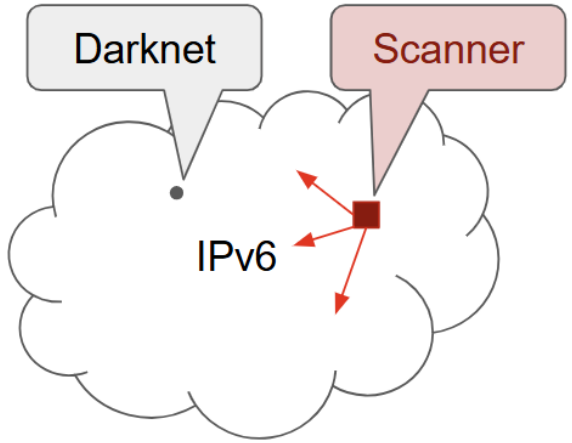
DarkBench framework

- Aim to unify the data, metrics, and methods to benchmark and perform comparable evaluation of algorithms that studies IBR
- We identified flaws in prior works



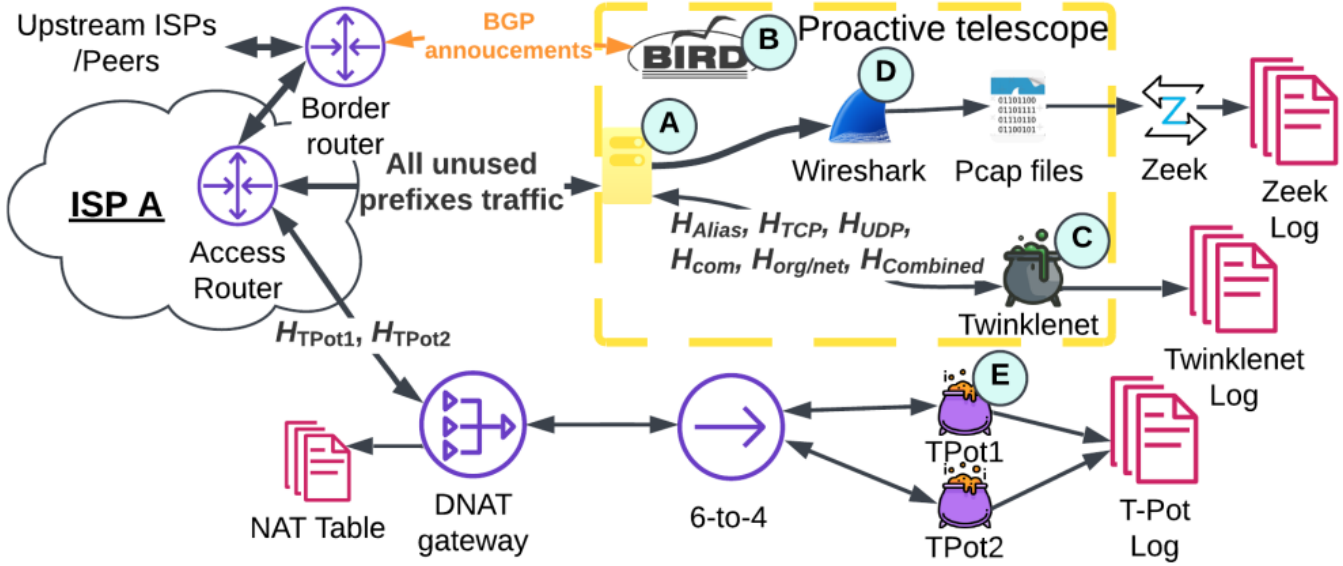
How about IPv6?

- Nowadays, we can scan the entire IPv4 in the order of hours
- Both passive and reactive approaches become ineffective in IPv6 network due to the vastness of IPv6 address space (2^{128} IP addresses)
  - Exhaustive scanning is practically impossible
  - Probing to random addresses have very low hit rate
  - The probability for darknets to observe these probe packets is slim
- Scanners changed their probing strategies in IPv6
  - Only probe the first address in each network
  - Employ IPv6 hitlists
  - Use Target Generation Algorithms (TGAs)
- ⇒ Discover “live” IPv6 hosts



Proactive network telescope

- We collaborate with a transit ISP (address space and infrastructure) to deploy a proactive network telescope
- BGP announcements
- Honeypots
- TLS Certificates
- Domain names
- Common subdomain names
- Addition to IPv6 Hitlists



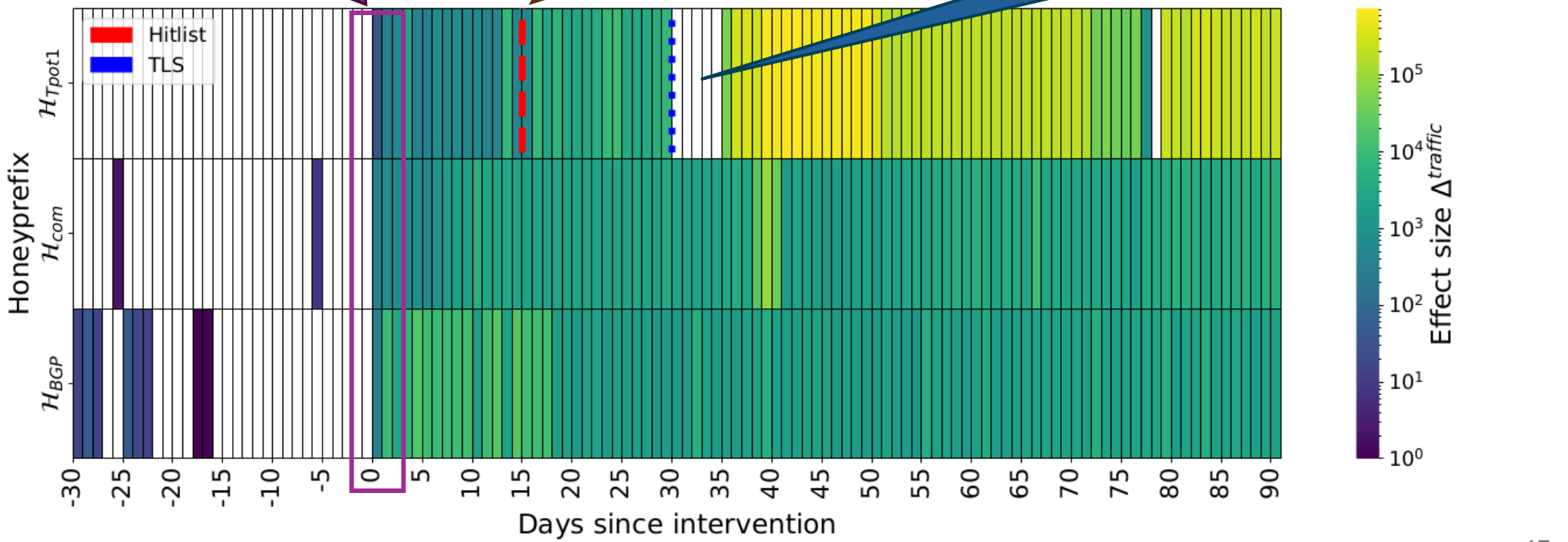
H. Tanveer et al. “Unveiling IPv6 Scanning Dynamics: A Longitudinal Study Using Large Scale Proactive and Passive IPv6 Telescopes”, In ACM CoNEXT 2025.

Effect of proactive features

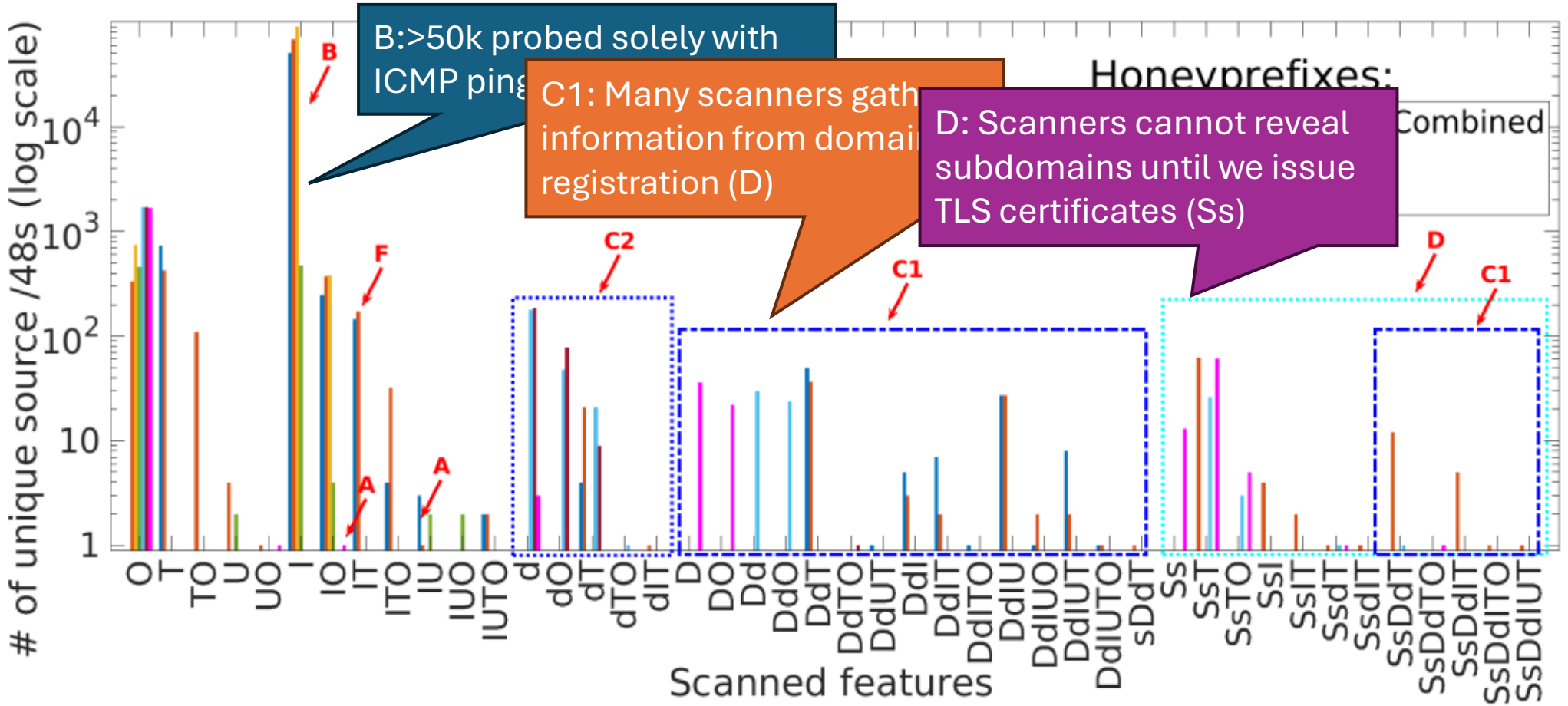
Initial BGP announcement receives immediate attention and long lasting

Inclusion in IPv6 hitlist triggers additional traffic

Issuing TLS certificates has strong effect size

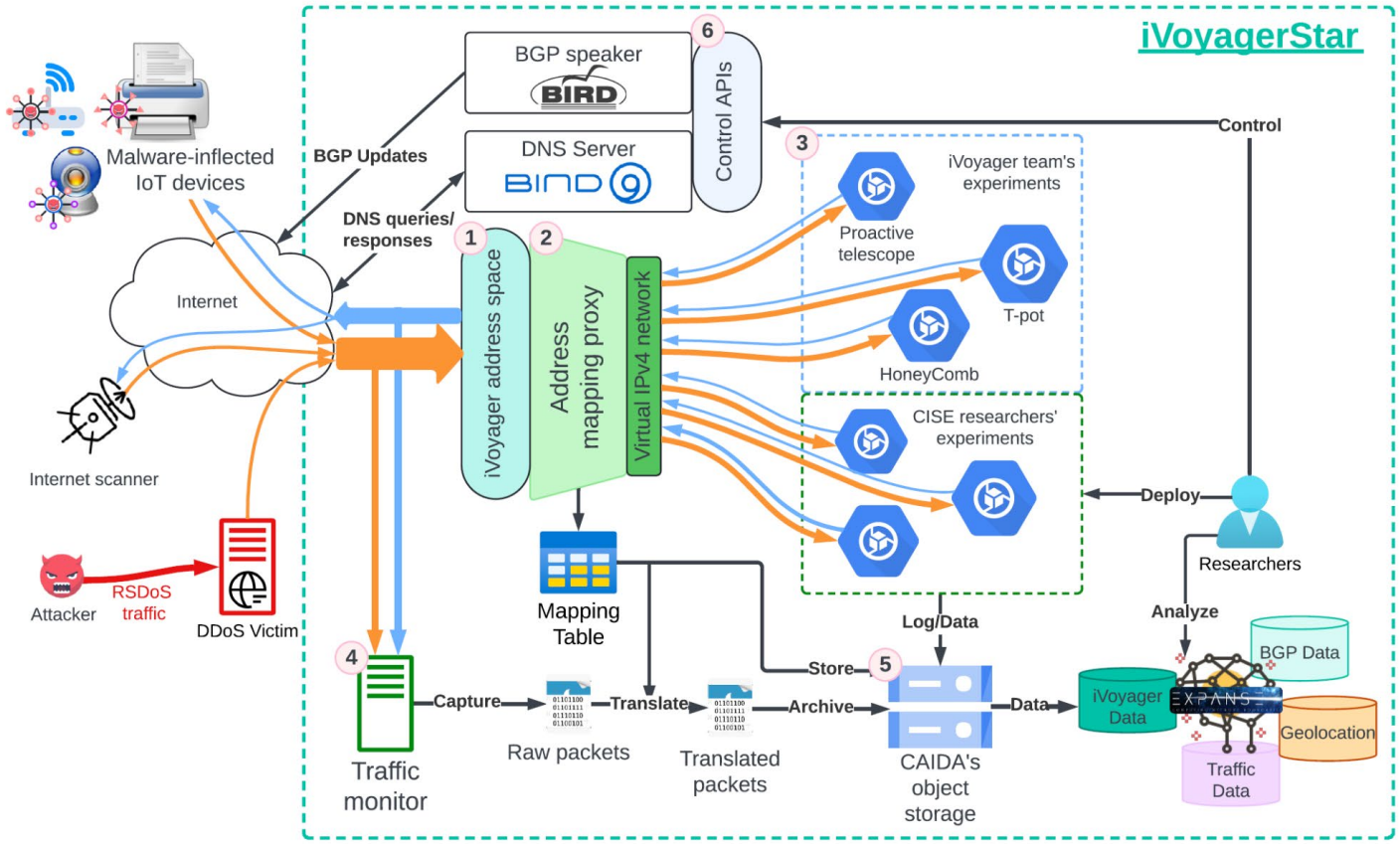


Scanning tactics



iVoyager

- Open IPv6 native research infrastructure for cybersecurity
- Lower the barrier in building infrastructure
- Provide data to evaluate TGAs



OAC-2450552

Conclusion

- Operational challenges of UCSD-NT
  - Root cause analysis and solutions
- Future directions in data dissemination of UCSD-NT
  - Facilitate data scientist to use the data
- Extend our monitoring capability to IPv6
  - iVoyager infrastructure for proactive network telescope implementation

Thank you

- Thanks to students at UCSD, collaborators at Akamai, ipinfo.io, MIT, Northwestern University, Louisiana State University, TU Dresden, University of Twente, University of Münster
- NSF awards # 2450552, 2531134, 2319959