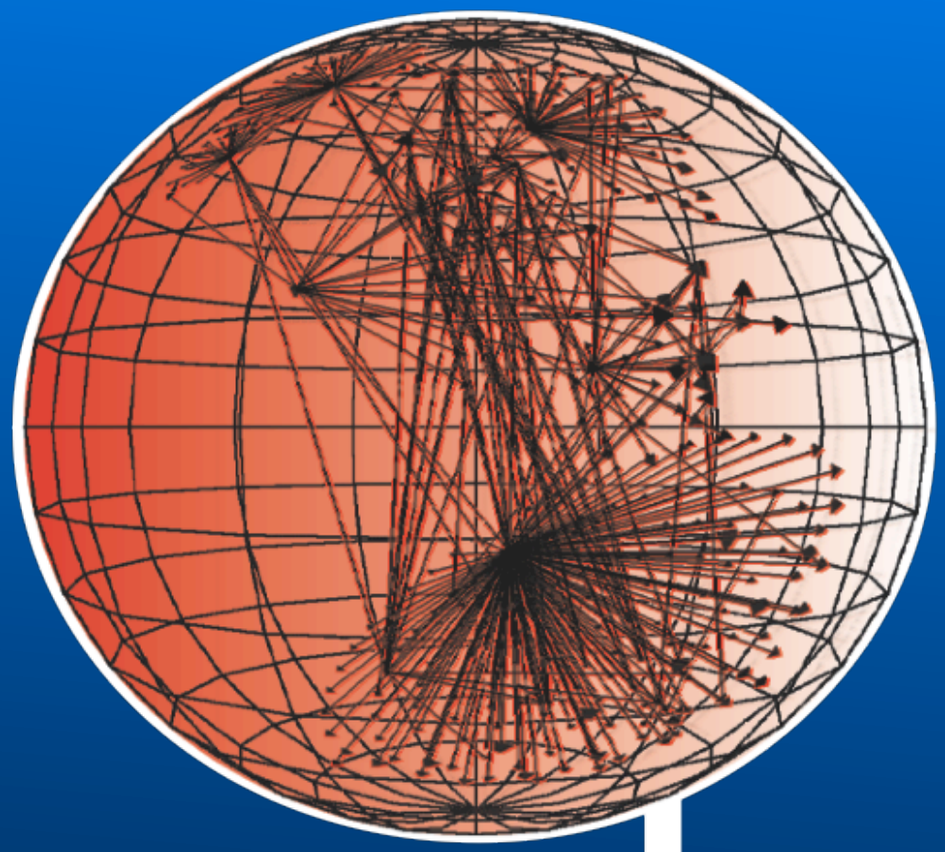
CAIDA Research Overview

kc claffy

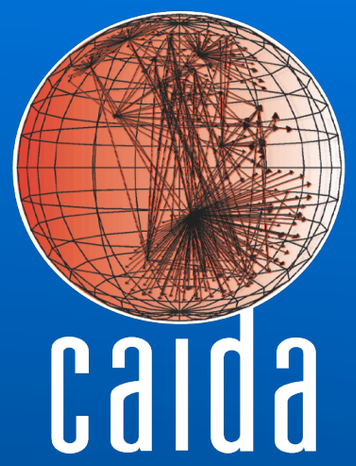
Spring 2013



caida

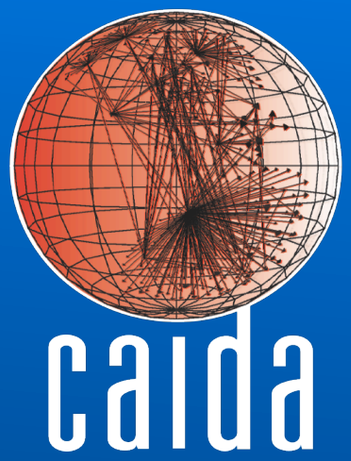
CAIDA's Mission Statement

CAIDA The Cooperative Association for Internet Data Analysis (CAIDA) is an independent analysis and research group based at the University of California's San Diego Supercomputer Center. CAIDA investigates both practical and theoretical aspects of the Internet.



Outline

- Introduction
- Measurement infrastructure / Data
- Research highlights
- Outreach



Outline

- Introduction
- Measurement infrastructure / Data
- Research highlights
- Outreach

CAIDA Measurement Infrastructure

- Archipelago (ark)
  - CAIDA active measurement infrastructure
  - supports ongoing topology measurement as well as customized experiments

- UCSD Internet Telescope (darknet)
  - packet capture to largely unused address space (one-way traffic only)

- Passive Trace Capture
  - captures packets on Tier 1 10GE backbone link (two-way traffic)
  - shared anonymized headers only

- HostDB
  - historical record of IP address to DNS hostname database

CAIDA Data Collections

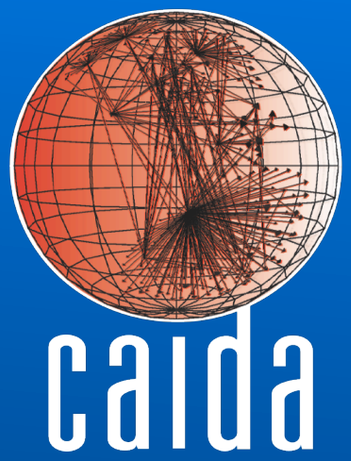
http://www.caida.org/data/overview/

- performance
  - DNS root/gTLD RTT DATA
- security
  - Computer Worms, Backscatter, DDoS attacks, Botnet scanning activity
- topology
  - AS Links, Prefix to AS, AS Rank, AS Relationships, IPv4+IPv6 topology, Macroscopic Internet Topology Data Kit (ITDK)
- traffic
  - Telescope Data, Telescope (live), Anonymized Internet Traces, Tier 1 packet traces, SDNAP
- DNS
  - historical record of IP address to DNS hostname database
- meta-data
  - DatCat

Research Highlights

http://www.caida.org/publications/

- topology analysis
  - Internet-scale router alias resolution
  - comparing IPv6 & IPv4 topology
  - Internet topology data sharing
- security & stability
  - large-scale Internet outages
  - botnet activity
- Internet peering analysis
  - inferring AS relationships
  - AS ranking
- interconnection economics
  - modeling peering strategies
  - transit pricing
- modeling complex networks
  - using hidden metric spaces
- geolocation analysis
  - comparing geolocation services
  - IP reputation vs. governance
- future Internet
  - IPv6
  - Named Data Networking
- visualization



Outline

- Introduction
- Measurement infrastructure / Data
- Research highlights
- Outreach

Archipelago

http://www.caida.org/projects/ark

- CAIDA's active measurement infrastructure
  - An environment for easy development and rapid prototyping of experiments.
- 71 monitors – growing 1 or 2 per month
  - 30 IPv6 capable
  - 33 countries
- current projects
  - team-probing experiment to collect IPv4 and IPv6 topology
  - alias resolution measurements
  - MIT ANA Spoofer project

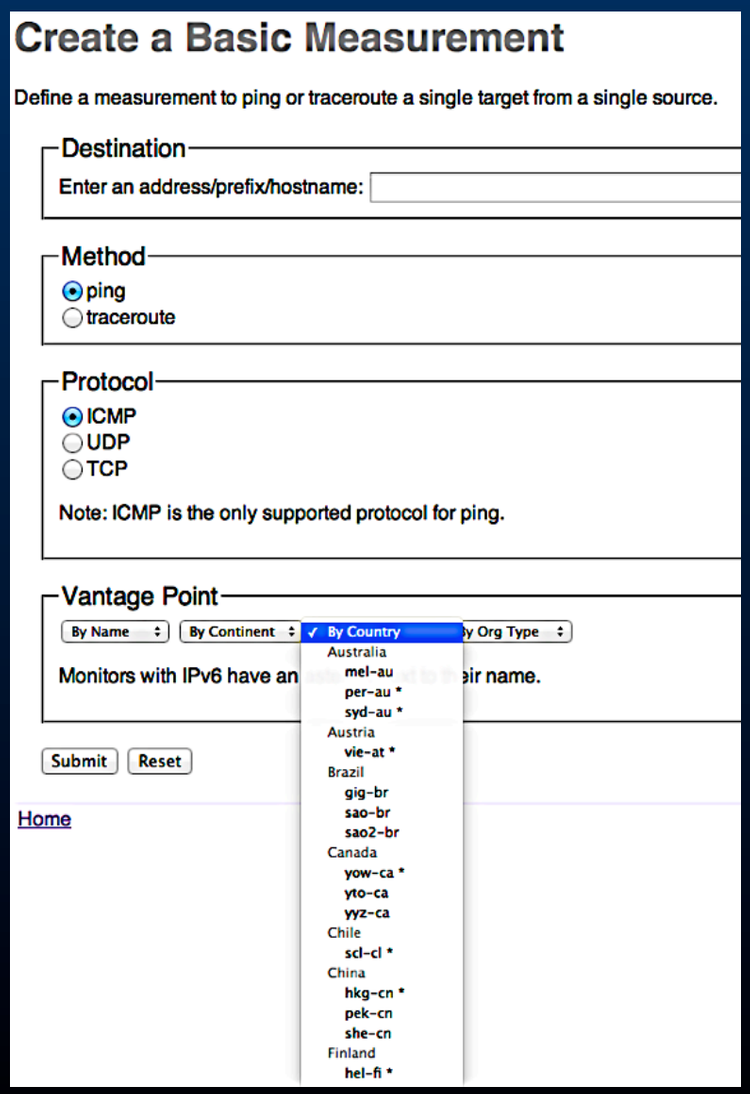
Archipelago

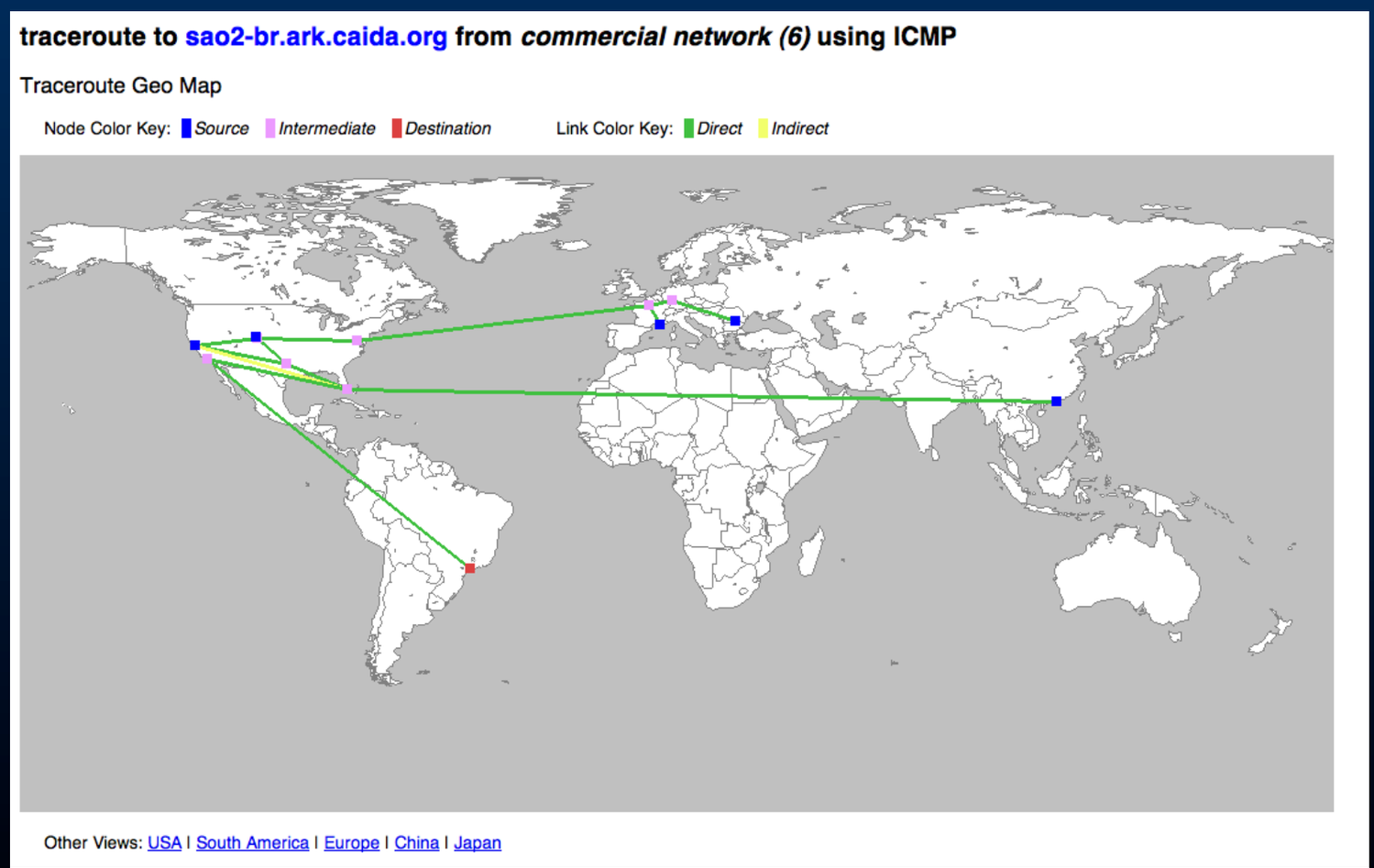
http://www.caida.org/projects/ark

caida

interactive web interface to **on-demand** topology **measurements** from Ark monitors

currently, ping and traceroute (ICMP, TCP, UDP)



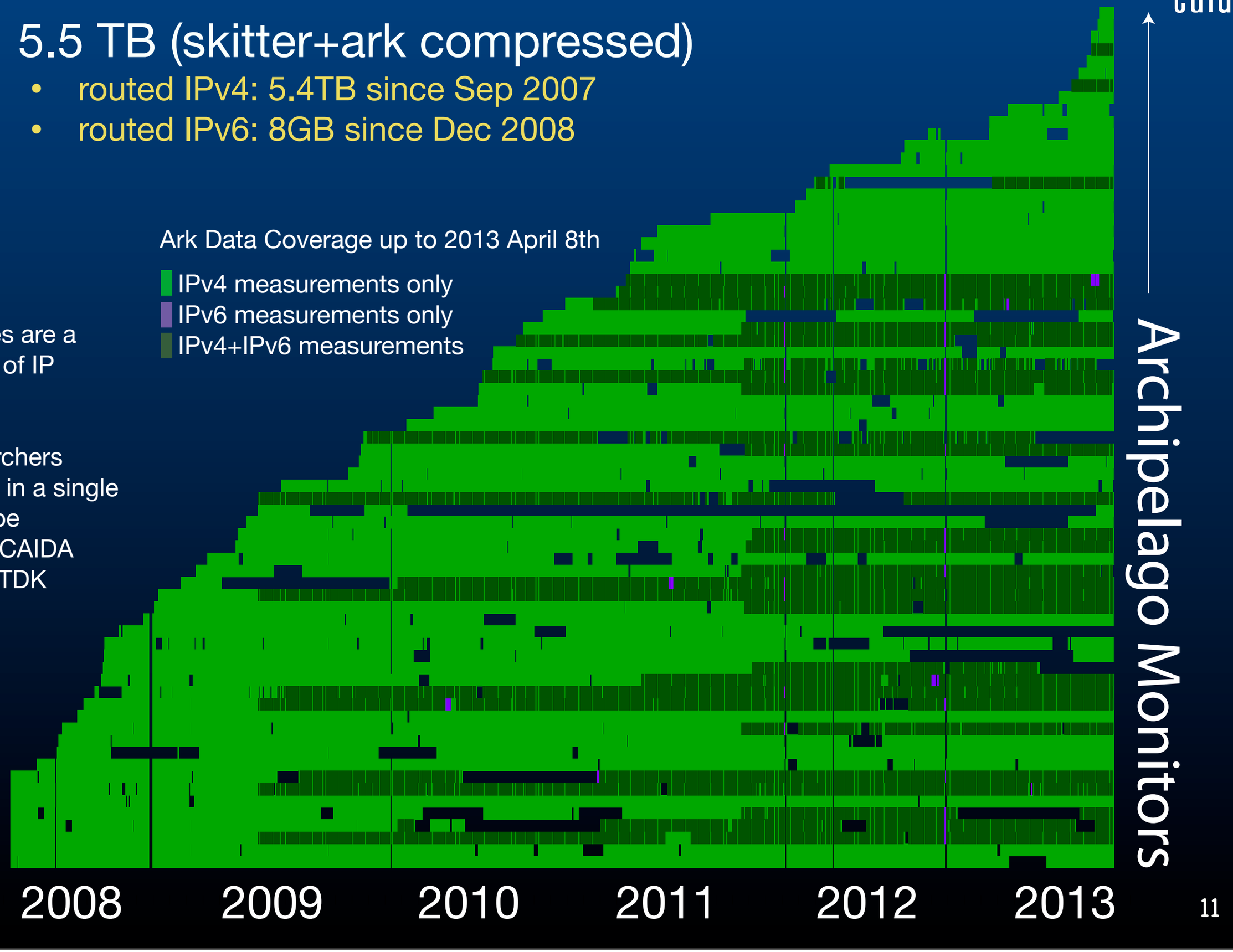


Archipelago monitors and data

- 5.5 TB (skitter+ark compressed)
  - routed IPv4: 5.4TB since Sep 2007
  - routed IPv6: 8GB since Dec 2008

Raw traces are a collection of IP paths

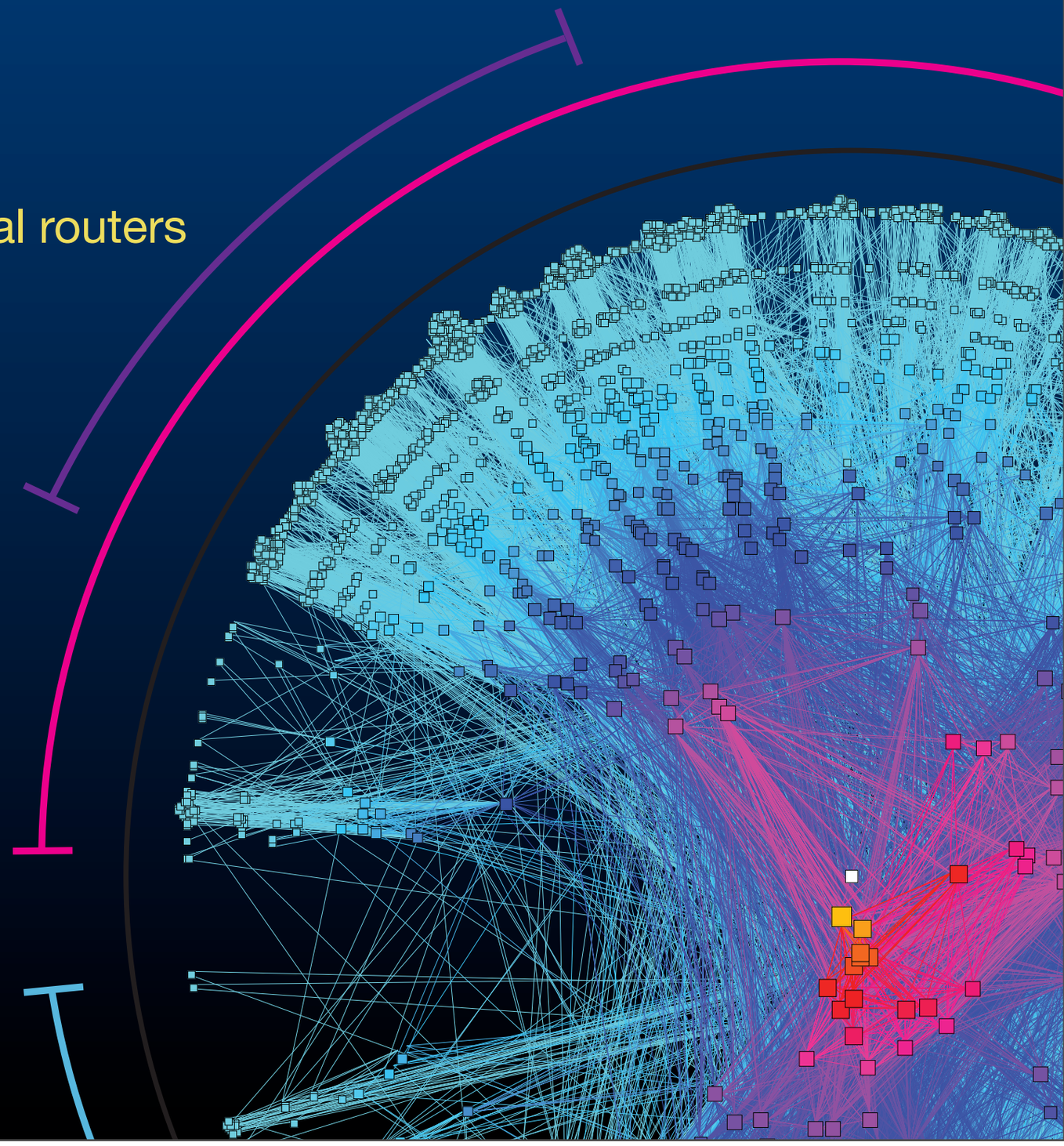
For researchers interested in a single microscope snapshot CAIDA provides ITDK



Internet Topology Data Kit (ITDK)

http://www.caida.org/data/active/internet-topology-data-kit/

- macroscopic Internet topology snapshot
  - provides a curated, annotated router level topology
- annotations
  - IP address aliased to routers
  - geographic location of individual routers
  - DNS hostnames for IP address
  - router to AS assignments
- multiple snapshots
  - Apr 2010
  - Jul 2010
  - Apr 2011



UCSD Network Telescope

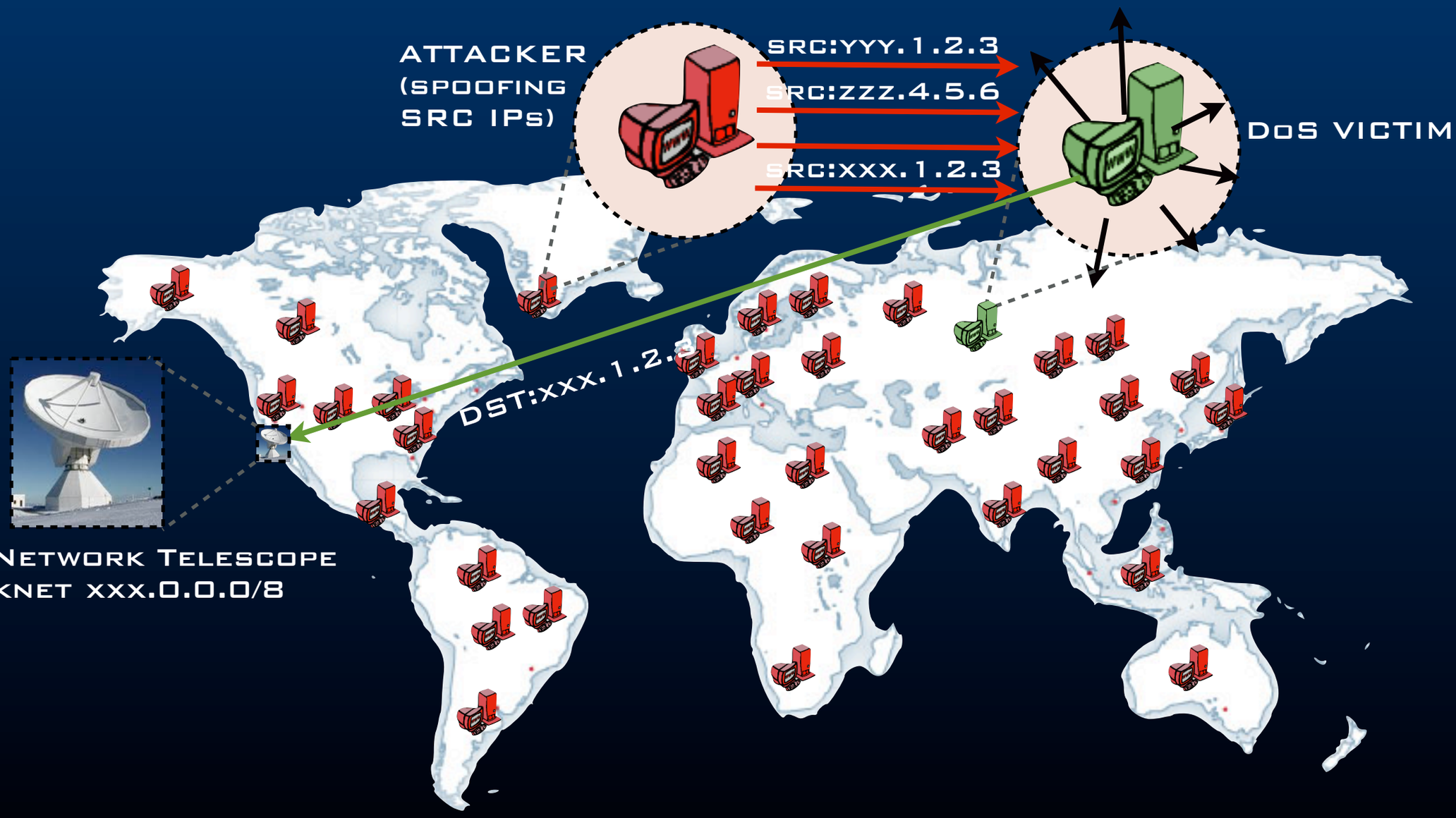
http://www.caida.org/data/passive/network_telescope.xml

- A portion of the Internet address space that is mostly unused - **1/256** of the **Internet** address space
- Receives *unsolicited* traffic (*Internet Background Radiation*)

UCSD Network Telescope

http://www.caida.org/data/passive/network_telescope.xml

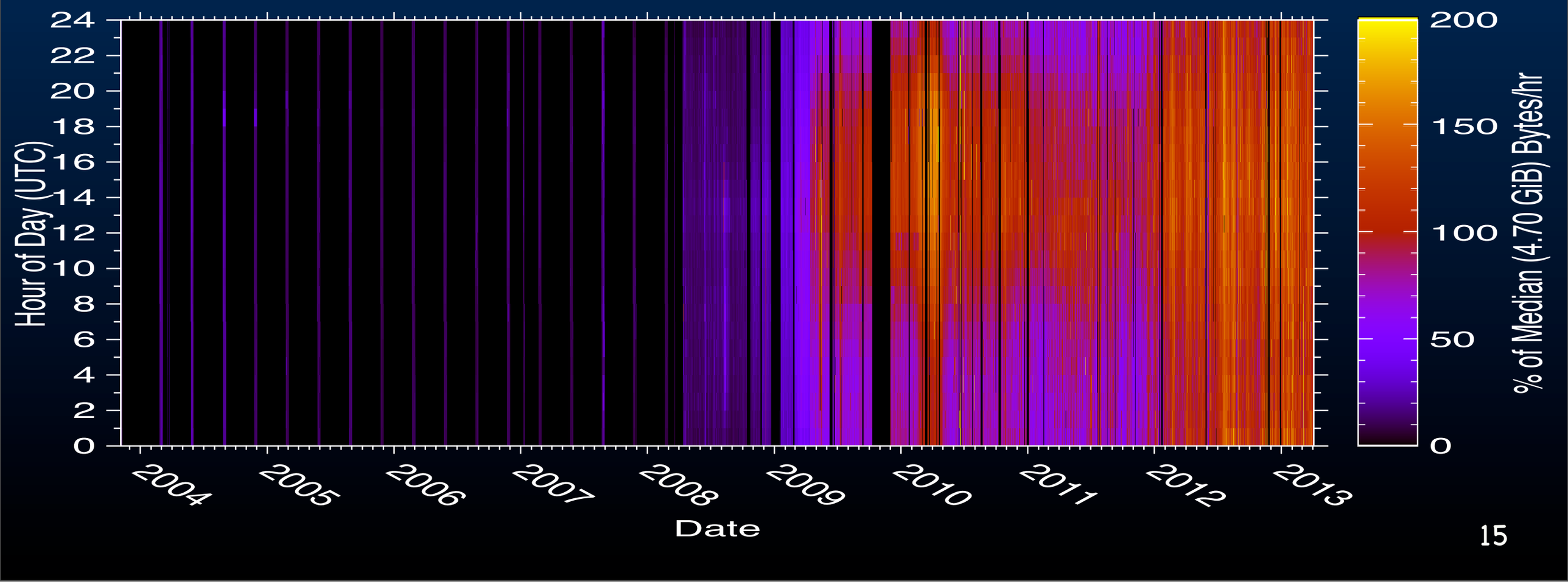
- part of IBR is *backscatter* from victims of randomly-spoofed *denial-of-service attacks*



UCSD Network Telescope

http://www.caida.org/data/passive/network_telescope.xml

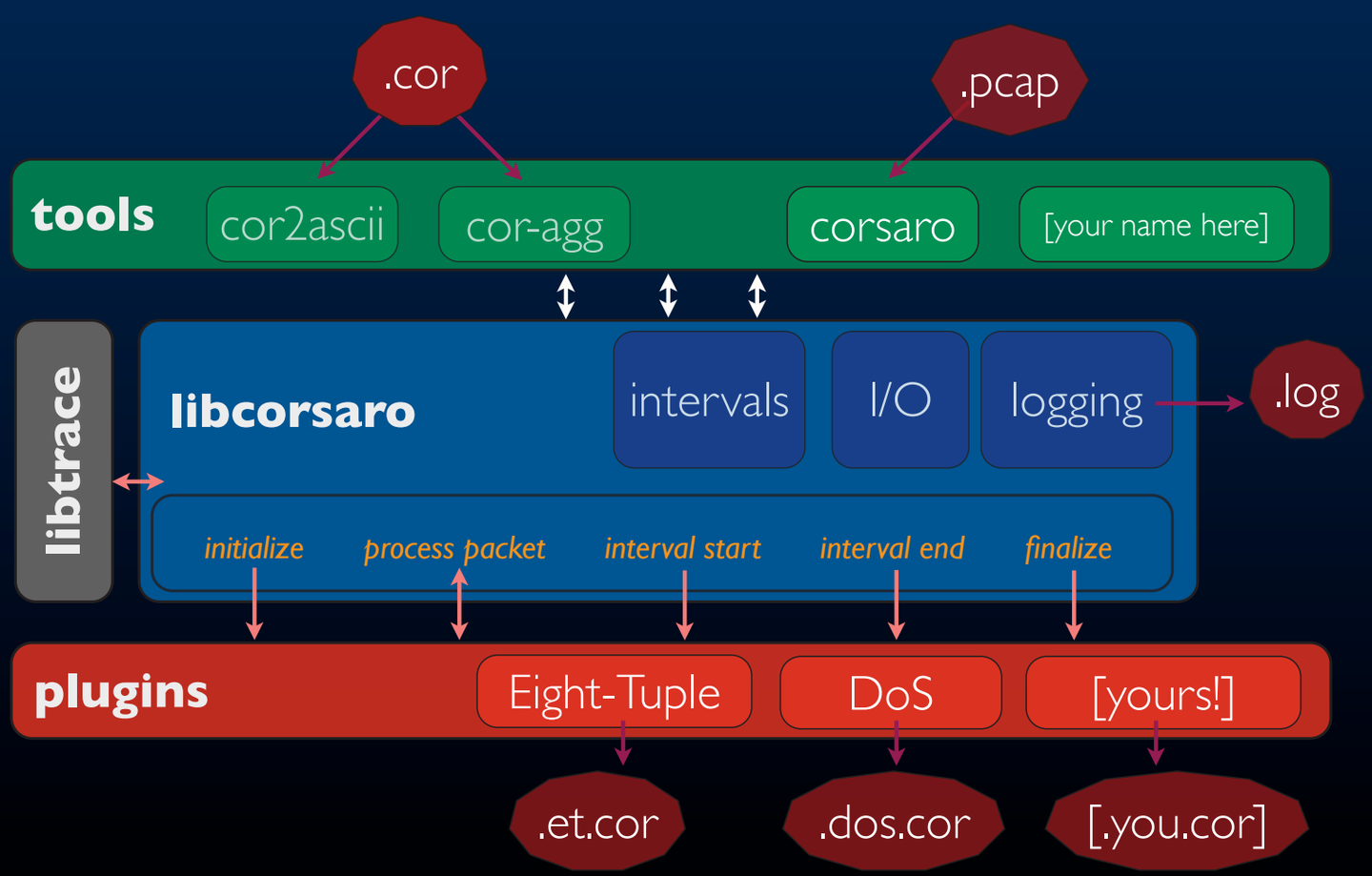
- packet capture of unsolicited traffic to unassigned addresses
- 2012 data still live on disk (currently ~146 days)
  - 16.92 TiB compressed, 34.28 TiB uncompressed
- data archived to NERSC in April
  - 105 TiB compressed/encrypted

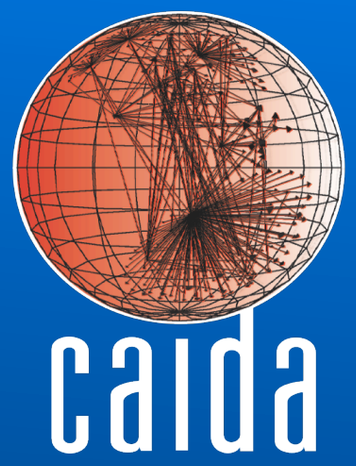


Corsaro - Darknet Traffic Analysis

http://www.caida.org/tools/measurement/corsaro/

- software framework for packet analysis, post-processing, data management
- aggregates data into intervals (e.g.,1-min bins)
- modular plugin-based architecture





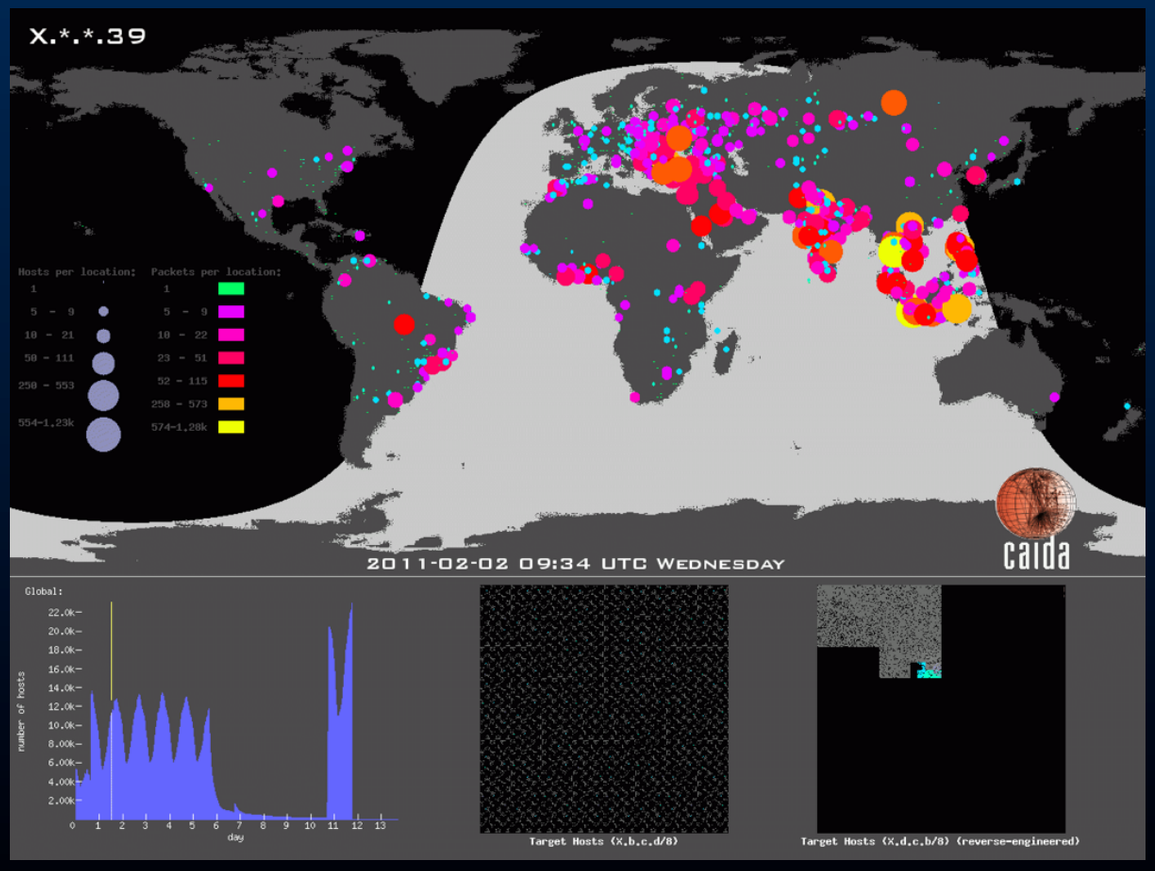
Outline

- Introduction
- Measurement infrastructure / Data
- Research highlights
- Outreach

Analysis of botnet stealth-scanning activity

A. Dainotti, A. King, kc Claffy, F. Papale, A. Pescapè

- a “/0” scan from a botnet
- observed by the UCSD Network Telescope
  - validated with multiple publicly available data sets
- scanning SIP servers with a specific query on UDP port 5060 and SYNs on TCP port 80



http://www.caida.org/publications/papers/2012/analysis_slash_zero/supplemental

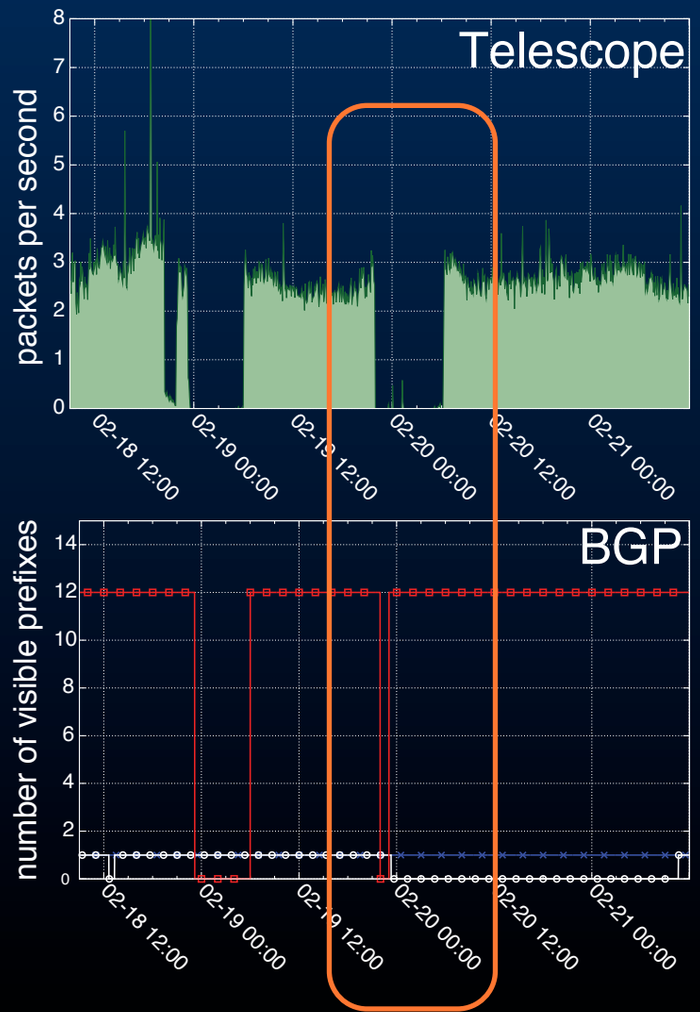
Large-scale Internet Outages

A. Dainotti, E. Aben, kc Claffy, A. King, K. Benson

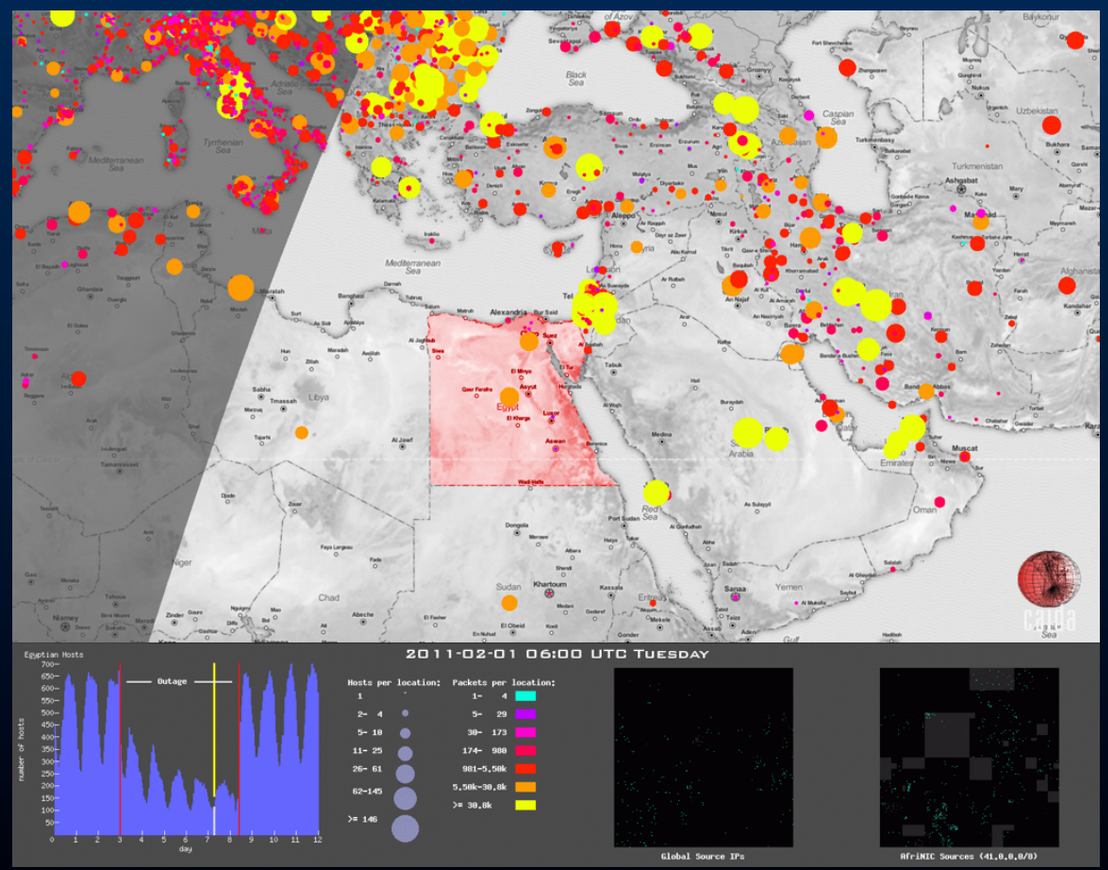
- combine BGP updates, Ark traceroutes, and darknet measurements to analyze country-wide outages during the “Arab Spring”

2012 Applied Networking Research Prize (IRTF)

Libya



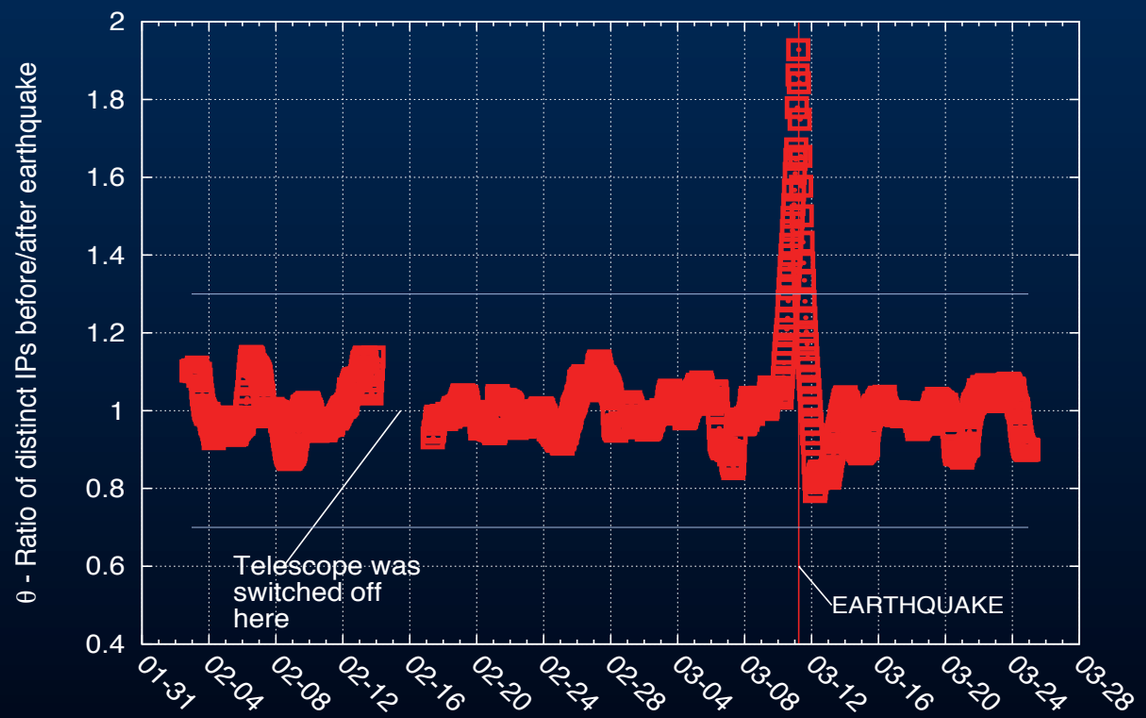
Egypt

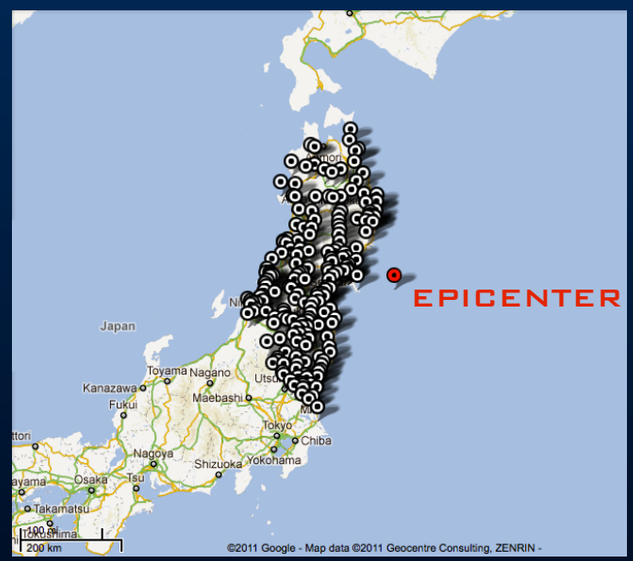


Large-scale Internet Outages

A. Dainotti, E. Aben, kc Claffy, A. King, K. Benson

- effects of natural disasters on the Internet infrastructure seen from the UCSD Network





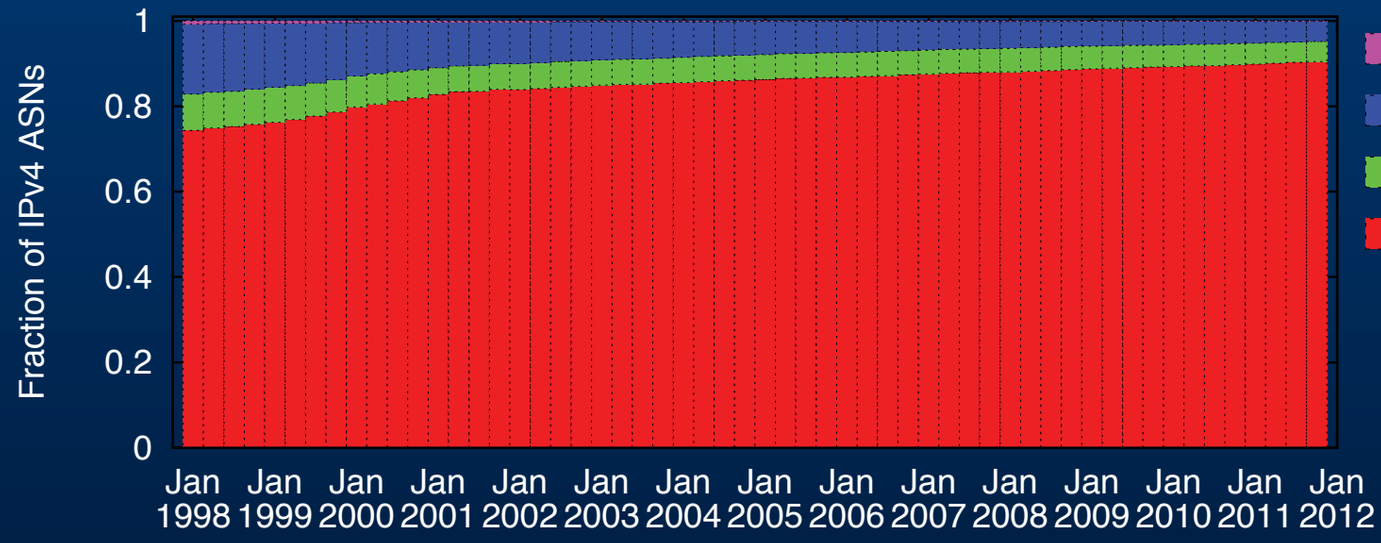
Japan, Mar 2011 Earthquake of Magnitude 9.0

Measuring the Deployment of IPv6

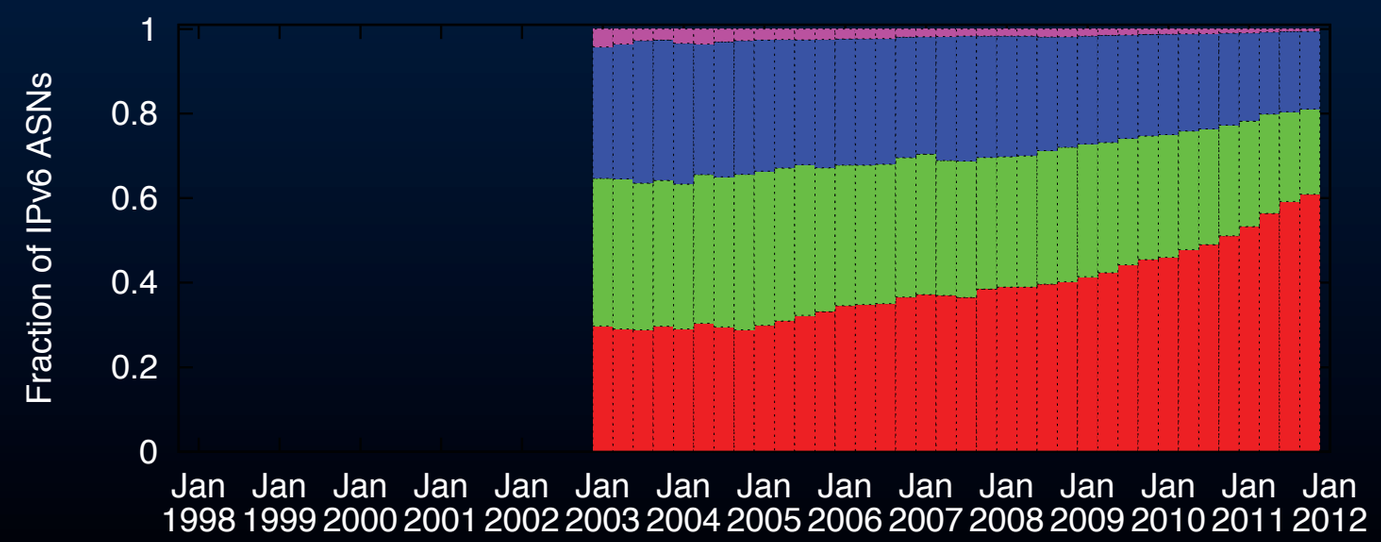
A. Dhamdhere, M. Luckie, B. Huffaker, kc Claffy, A. Elmokashfi, E. Aben

caida

IPv4



IPv6

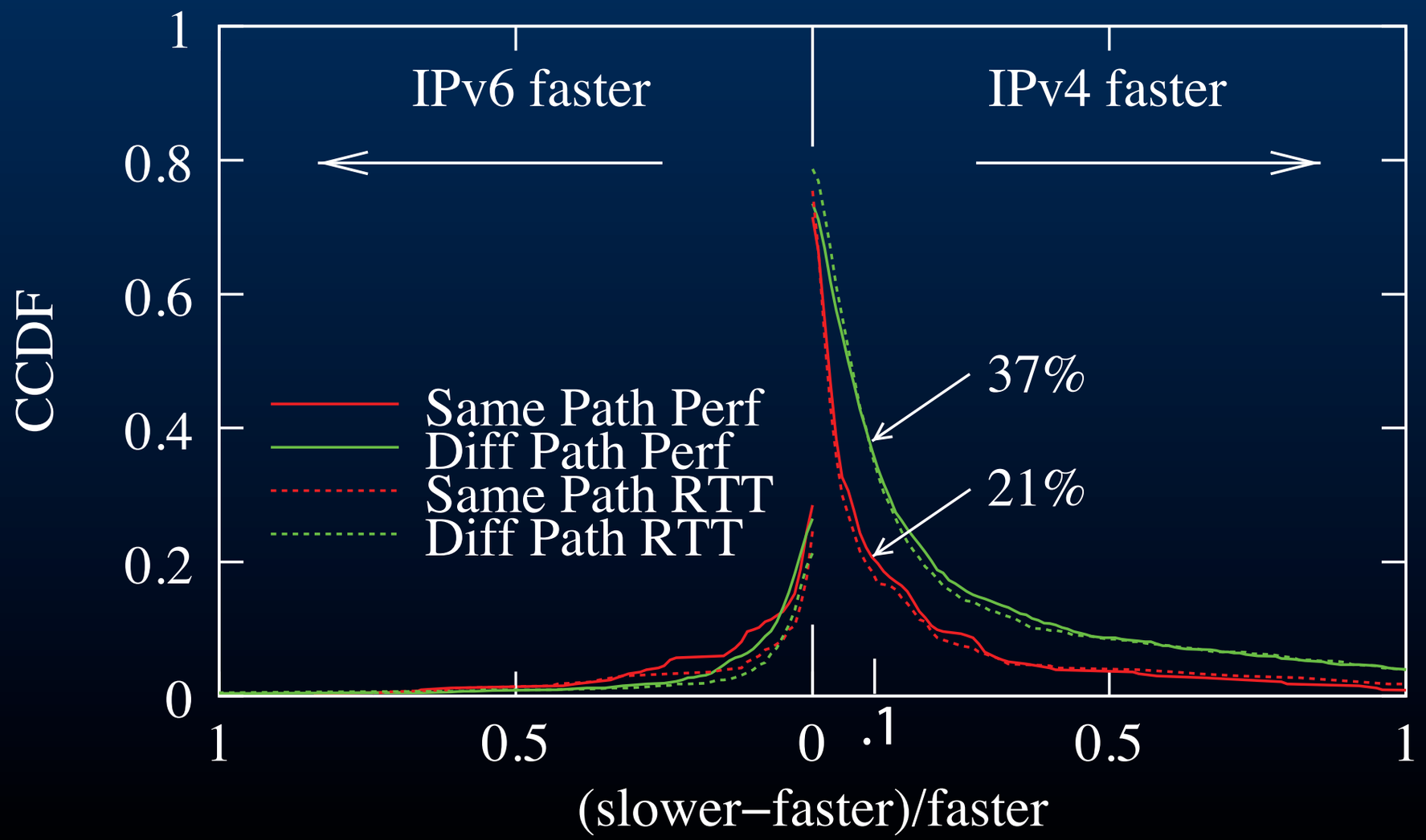


Breakdown of ASes by type. Over time IPv4 is becoming more like IPv6.

Measuring the Deployment of IPv6

A. Dhamdhere, M. Luckie, B. Huffaker, kc Claffy, A. Elmokashfi, E. Aben

- Mean RTT vs Performance
  - Relative performance between IPv4 and IPv6 measured by the relative mean fetch times (solid lines) and minimum SYN/ACK RTT (dashed lines).
  - When the IPv4 and IPv6 paths where the same (red) and different (green)

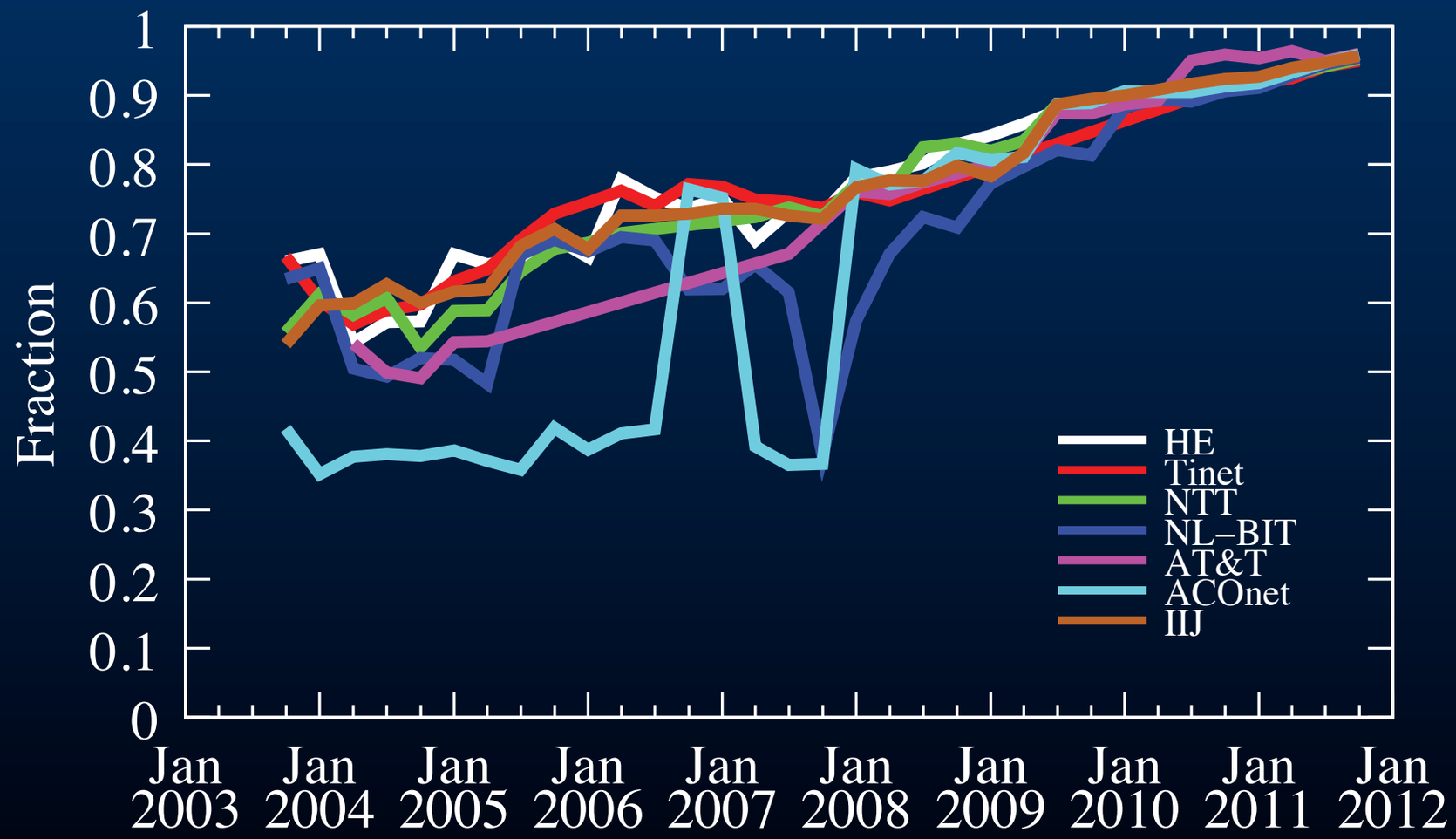


Measuring the Deployment of IPv6

A. Dhamdhere, M. Luckie, B. Huffaker, kc Claffy, A. Elmokashfi, E. Aben

IPv6 Paths that Could Be (but aren't yet)

http://blog.caida.org/best_available_data/2012/06/04/ipv6-what-could-be-but-isnt-yet/

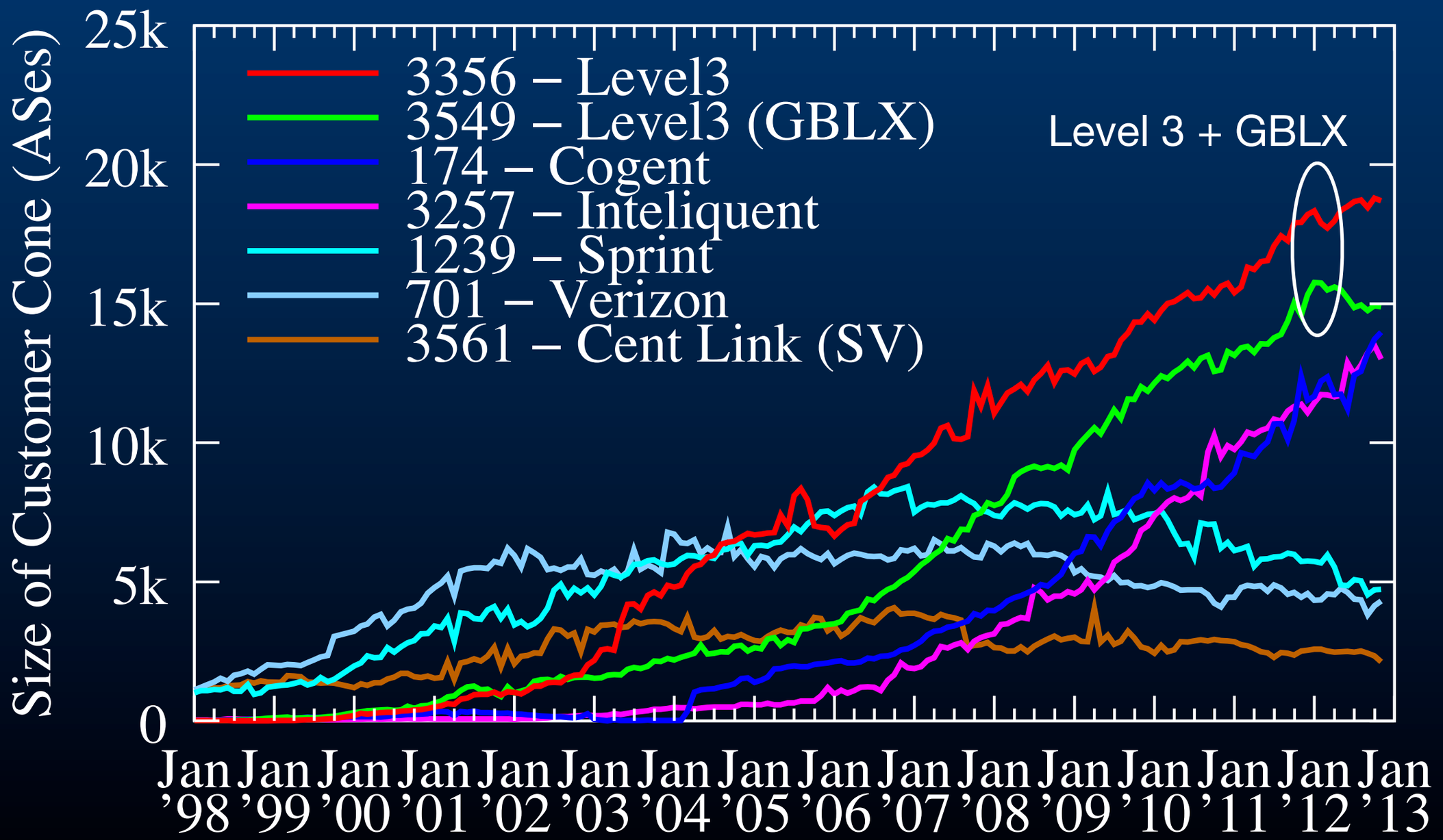


fraction of dual stack paths that could be node identical

AS Relationships, Customer Cones, and Validation

M. Luckie, B. Huffaker, A. Dhamdhere, V. Giotsas, kc Claffy

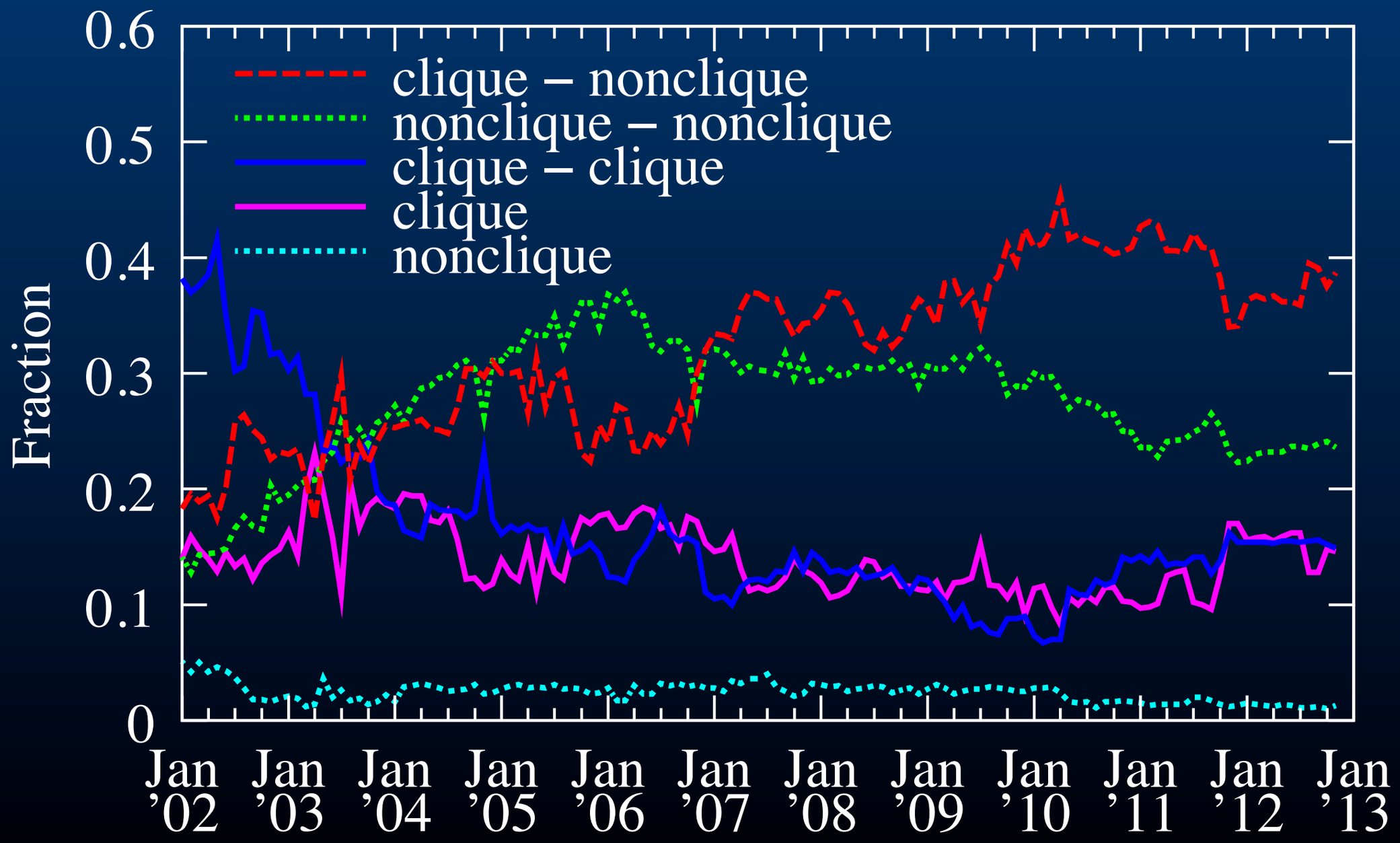
- Customer Cone (ASes) over time for largest transit providers



AS Relationships, Customer Cones, and Validation

M. Luckie, B. Huffaker, A. Dhamdhere, V. Giotsas, kc Claffy

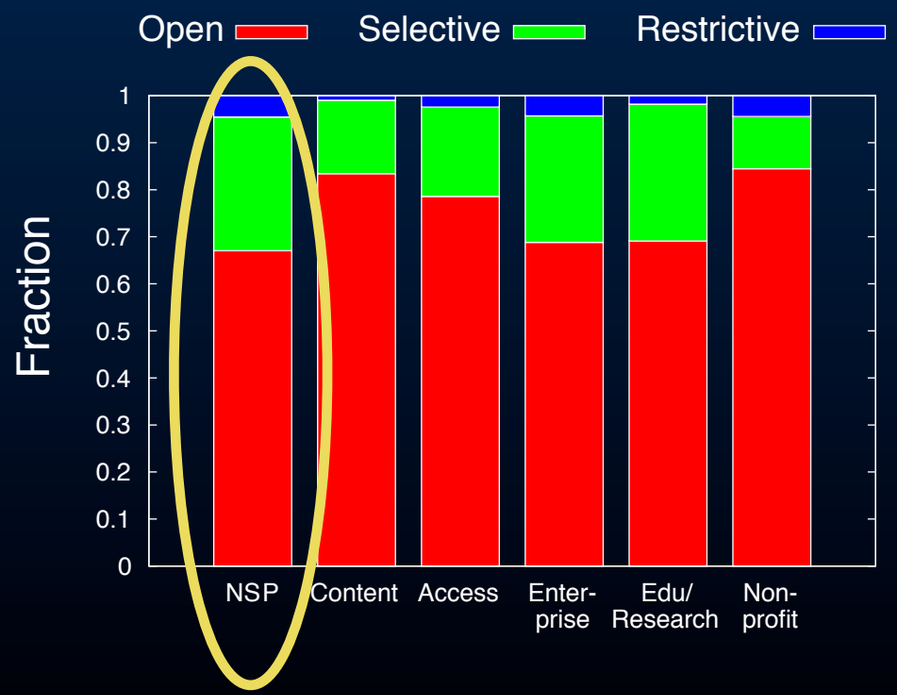
- Changes in Top of AS paths

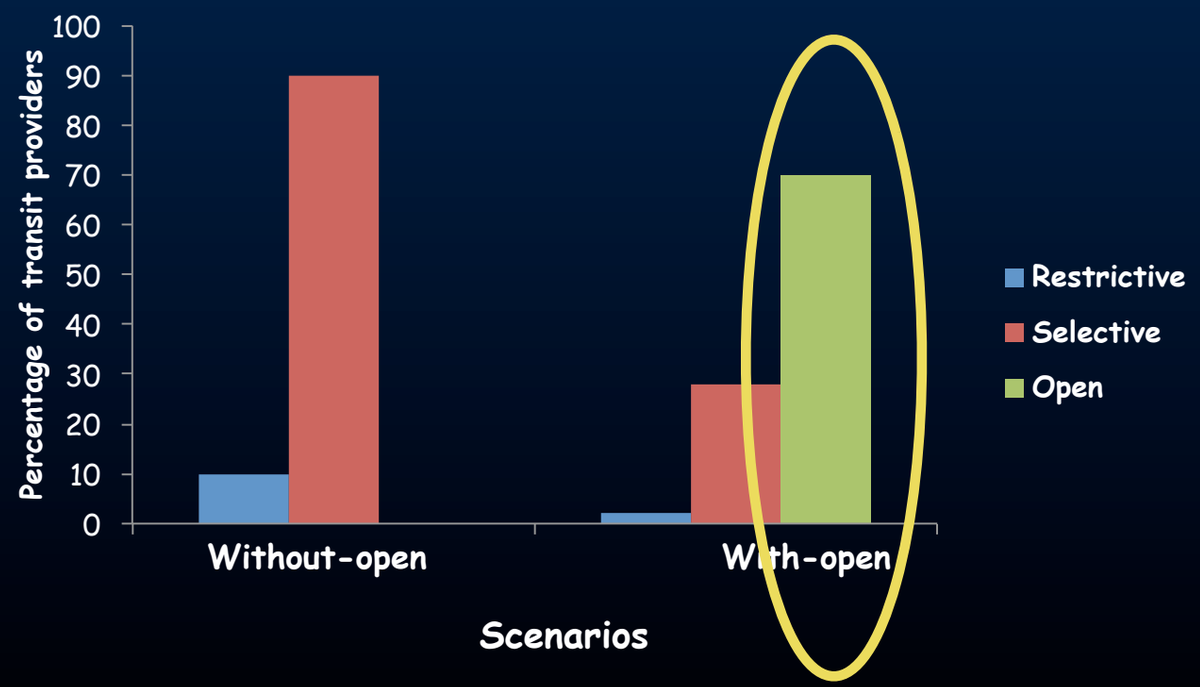


Modeling Peering Strategy by Transit Providers

A. Lodhi, A. Dhamdhere, C. Dovrolis

- Most transit networks advertise open peering in peeringDB.com
- Game-theoretic analysis shows that myopic, decentralized decisions cause gravitation toward open peering
- Simulations incorporating geography, traffic, peering, costs, & revenue reproduce real-world gravitation to open peering, *predict loss of economic fitness for many transit providers*





Transit Pricing Methods

V. Reddy, A. Dhamdhere, S. Shakkotai, S. Leinen, kc Claffy

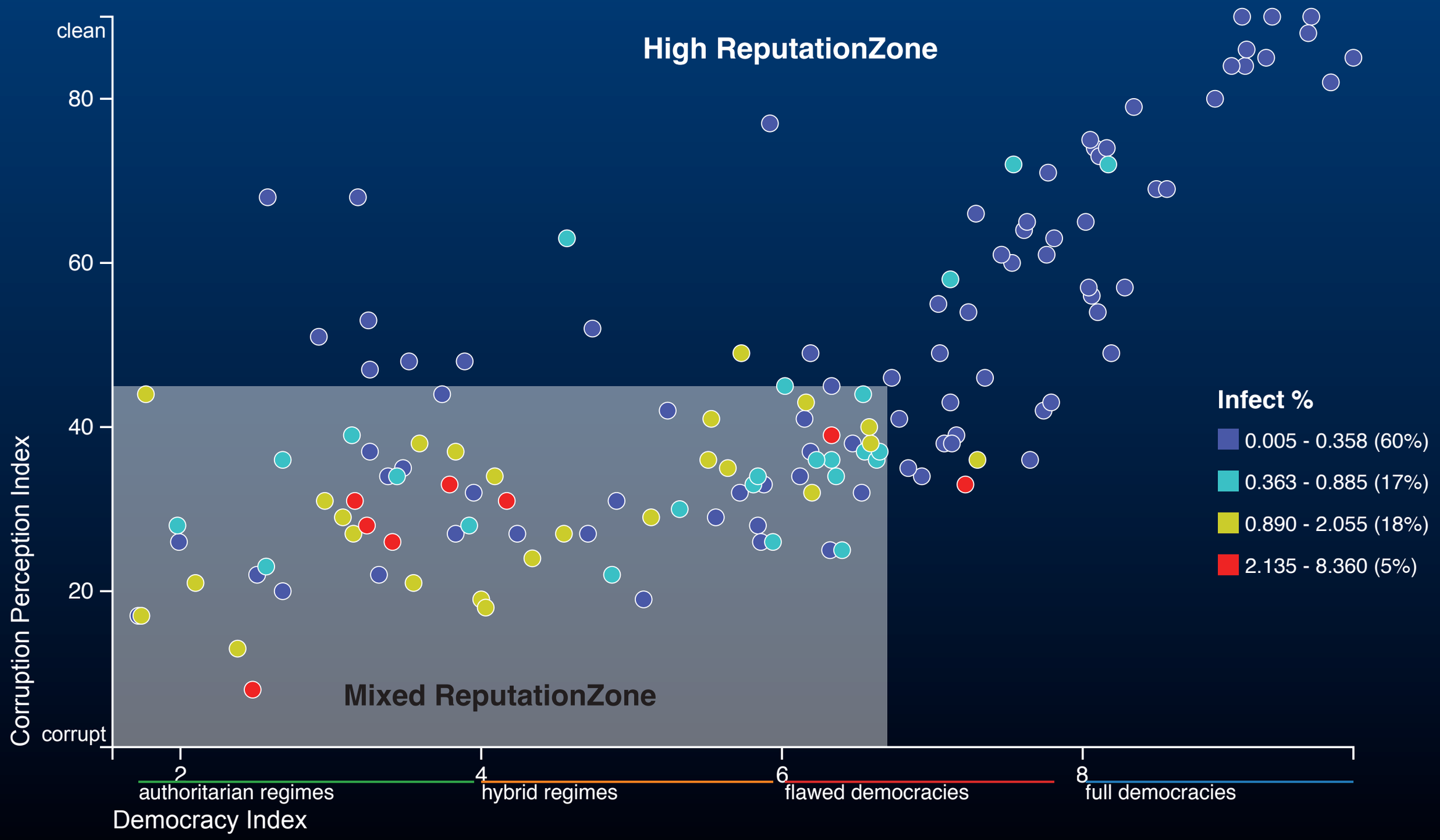
ongoing Cisco-funded work

- Transit providers have used 95th percentile pricing for more than a decade
- Is 95th percentile a rational choice for all customers?
- **Part 1:** Study historic traffic patterns between networks
  - interdomain traffic data from SWITCH
  - measure relation between 95th percentile and avg/peak over time
  - analyze for different traffic types and customer types
- **Part 2**: Optimize charging percentiles per-customer.
  - Each customer does not need to be charged at the same percentile!
  - Transit provider can set price to better match costs
  - Can offer discounts (lower charging percentiles) to low-cost customers based on traffic profile

Country level IP Reputation

B. Huffaker, kc Claffy

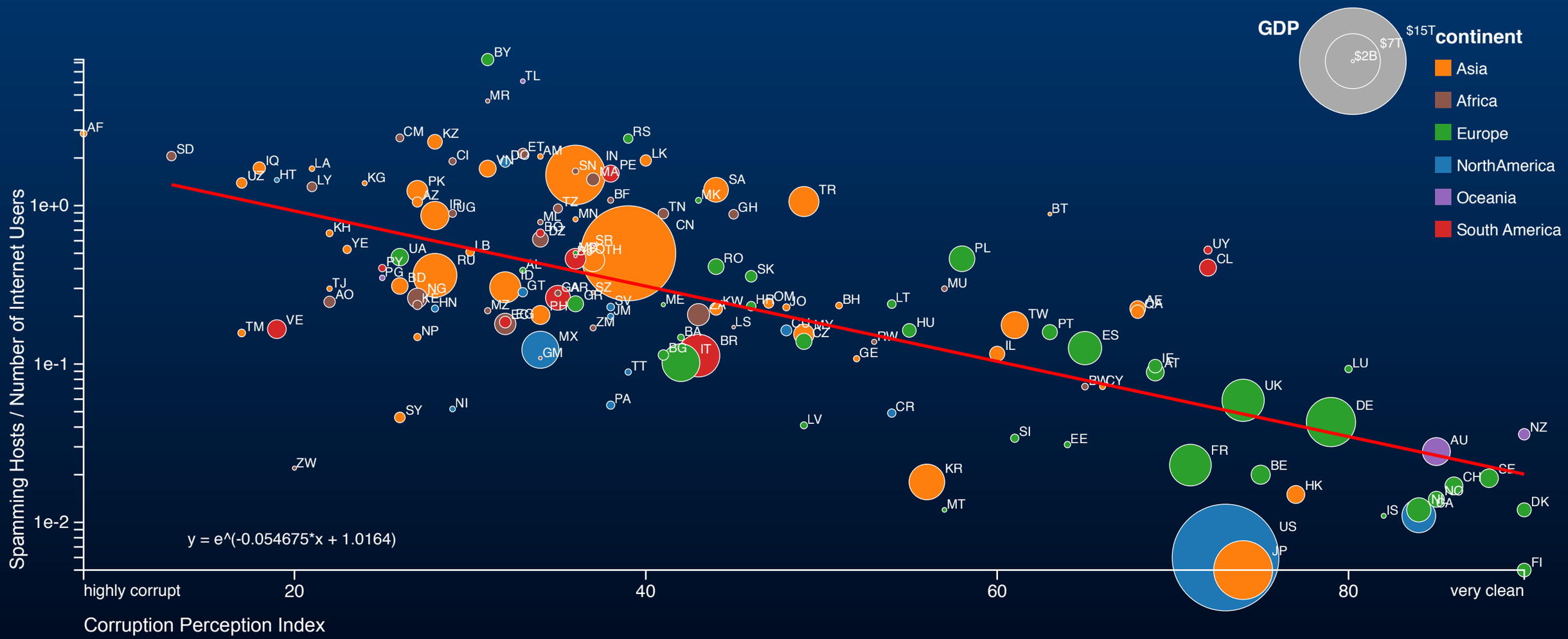
caida



Country level IP Reputation

B. Huffaker, kc Claffy

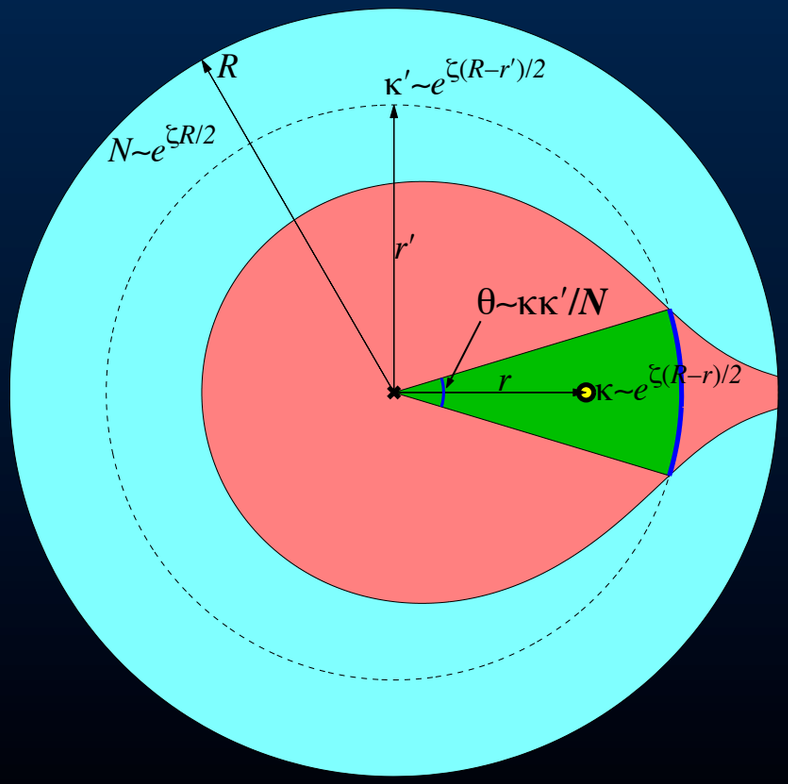
caida

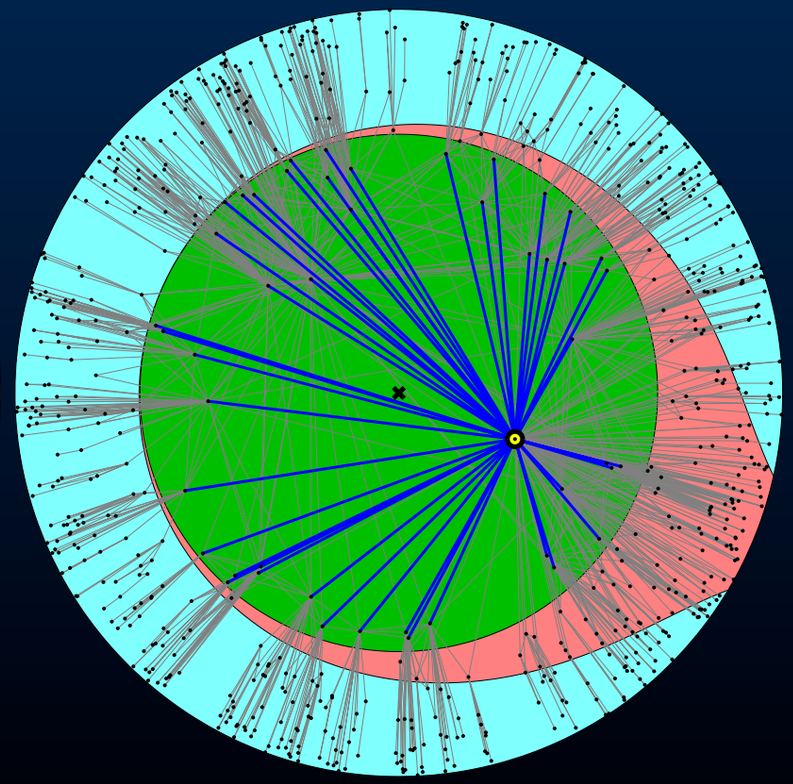


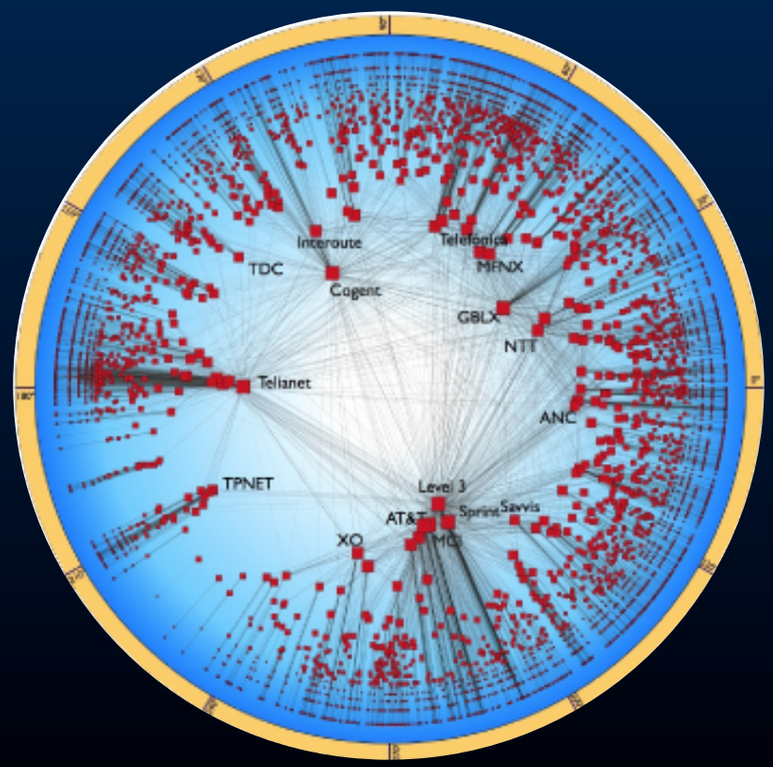
Modeling Complex Networks

D. Krioukov, M. Kitsak

- hidden variables influence structure of network
  - (which nodes are connected)
- variables form a hidden metric space that can be used to enable shortest path routing without global knowledge of topology
  - suggests potential direction toward infinitely scalable Internet routing architecture



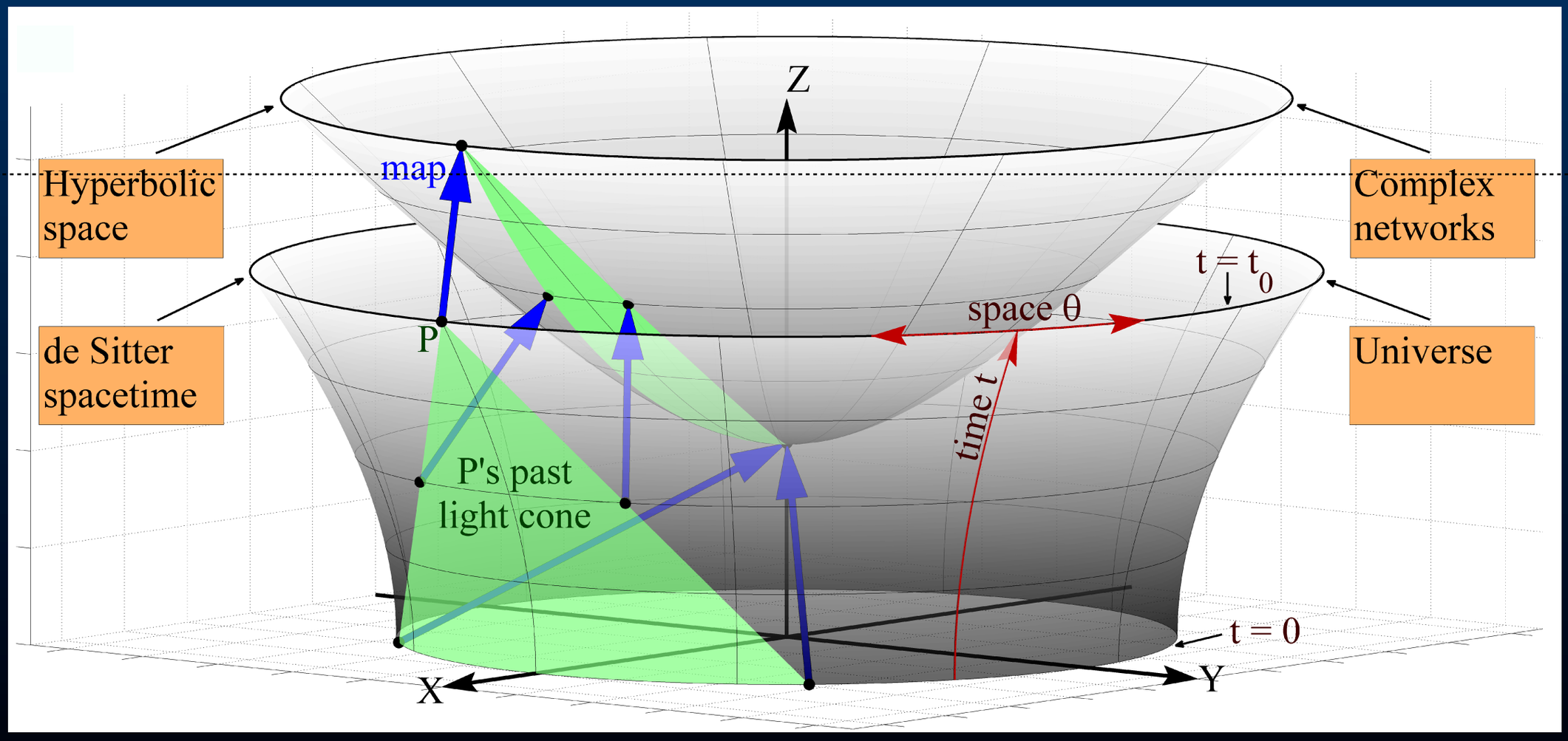




Network Cosmology

D. Krioukov, M. Kitsak, R. Sinkovits, D. Rideout, D. Meyer, M. Boguna

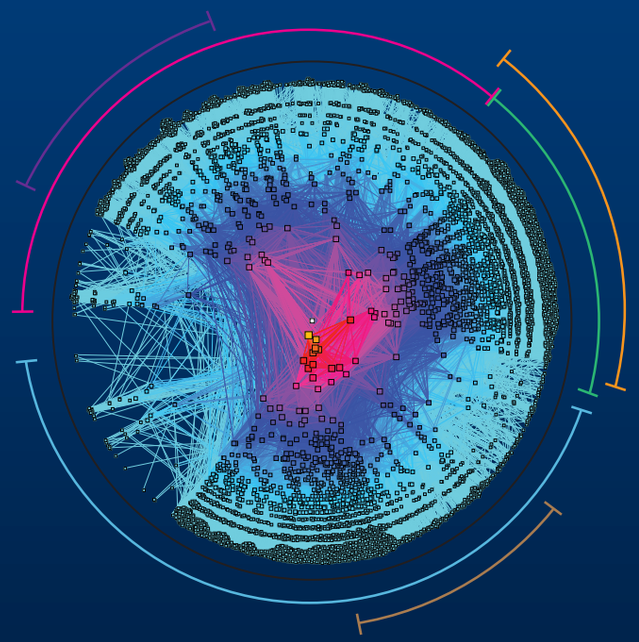
Representation of causal network of spacetime is similar to complex networks such as the Internet, social, or biological networks.



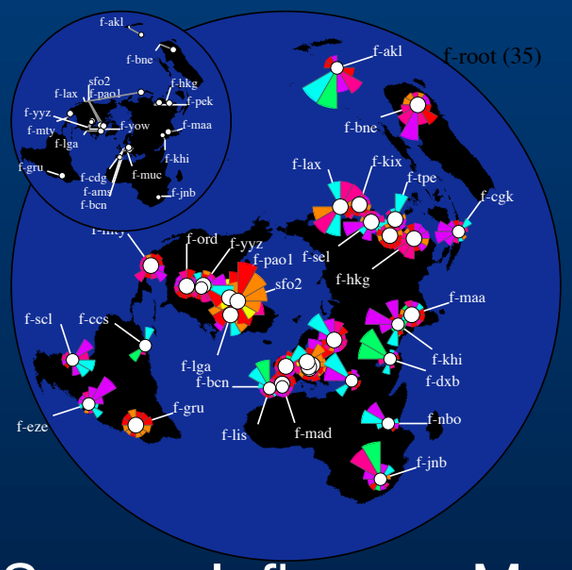
Visualization

B. Huffaker, A. King, Y. Hyun, A. Dainotti

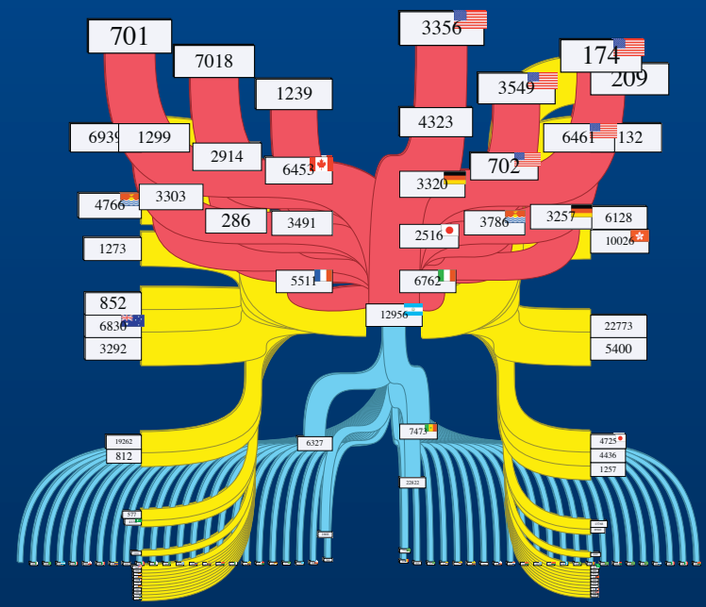
caida



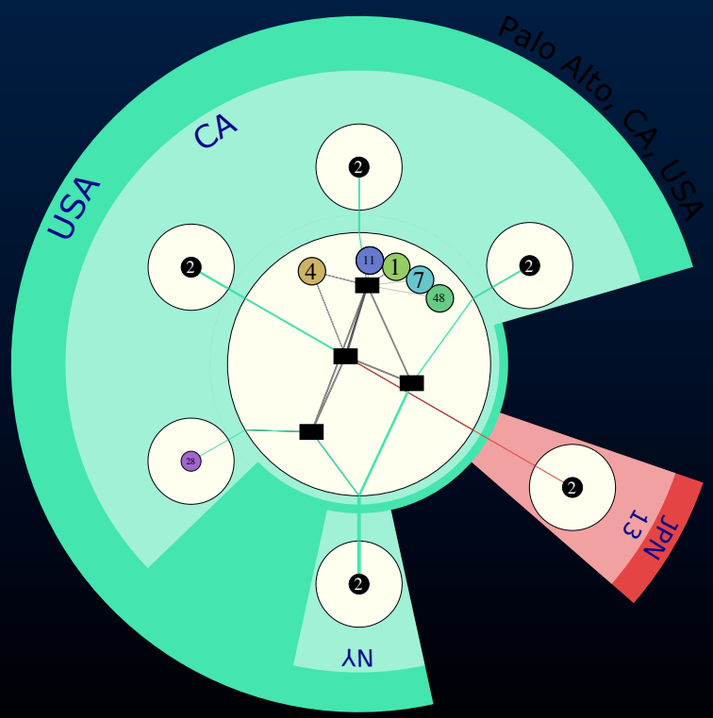
AS Core 2011



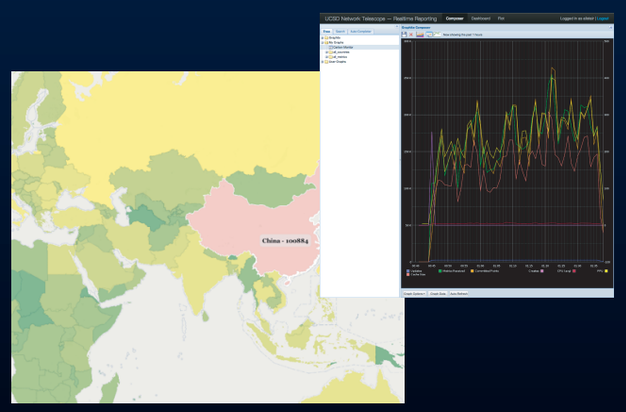
Source Influence Map
F-root Instances



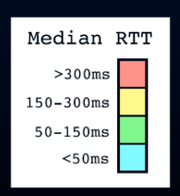
AS Relationships for
Telefonica (12956)



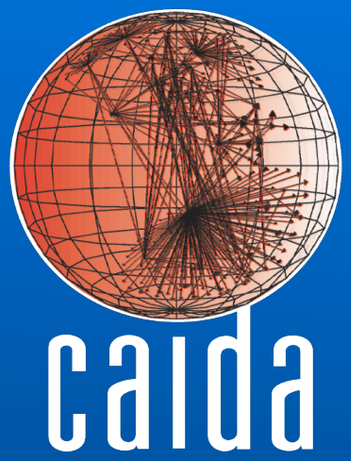
AS Pop Level



Coordinated/Interactive
views with cartography



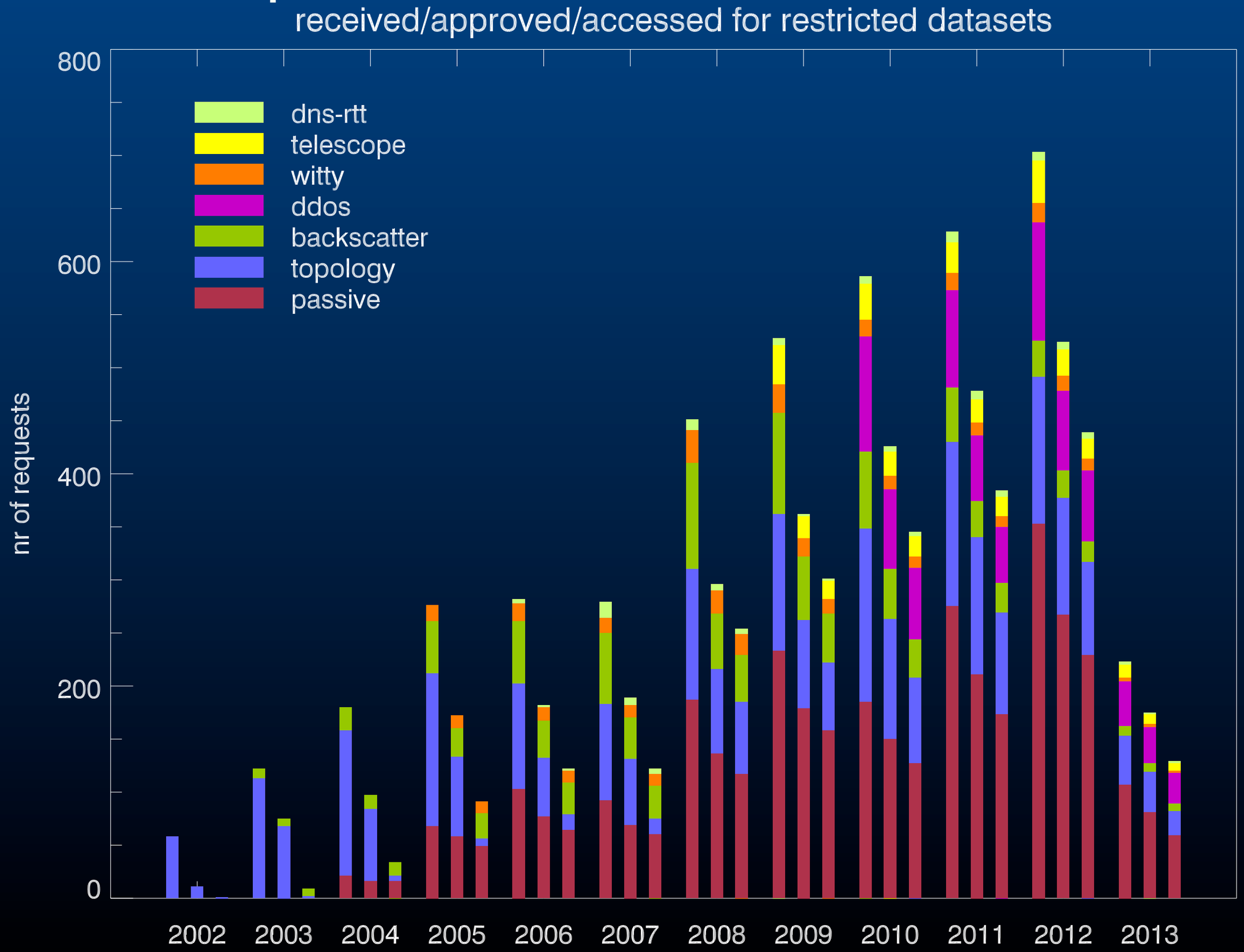
RTT per hop from
A-root monitor



Outline

- Introduction
- Measurement infrastructure / Data
- Research highlights
- Outreach

Data Requests



Data Set Popularity Ranking 2011~2013

- 1st: IP Packet Header traces (10GE/OC192, OC48)
  - 735 requests (559 approved, 461 accessed data)
  - who used it: 282 .us (.edu+.com+mil+.gov), 110 .cn, 45 .uk, 20 .ca, 15 .jp
    - 62 domains
    - 1244 total accounts: 270 from U.S.
- 2nd: topology data
  - 339 requests (277 approved, 207 accessed data)
  - who used it: 144 .us (.edu+.com+.mil+.gov), 55 .cn, 19.uk, 12 .kr, 6 .jp
    - 46 more domains
    - 892 total accounts

Publications using CAIDA data

http://www.caida.org/data/publications/bydataset/index.xml

- OC192 and OC48 traces: *traffic classification, performance modeling, monitoring, filtering, generation, locality*
  - **172 publications - *Mar 2013***
  - 82 publications - *Jan 2011*
- UCSD Network Telescope: *Conficker, worms, anomaly detection, denial-of-service attacks*
  - **43 publications - *Mar 2013***
  - 11 publications - *Jan 2011*
- Topology: *pkt traceback, marking, DOS defense, topology and routing modeling, discovery, metrics, methodology improvements*
  - **250 publications - *Mar 2013***
  - 108 publications - *Jan 2011*

Data availability

PREDICT: OC48 traces, topology, telescope

Derived data sets publicly available (i.e., AS-links)

- sample use: *http://semilattice.net/projects/map-of-the-internet/*

Academics who sign updated AUA

http://www.caida.org/data/

Commercial researchers

- a small sample of ‘industry evaluation’ data to entice interest
  join CAIDA, various membership levels are offered

Workshops

http://www.caida.org/workshops/

- Active Internet Measurement Systems (AIMS)
  - Feb ’09, Feb ’10, Feb ’11, Feb ’12, Feb ’13
- BGP/Traceroute Workshop
  - Aug ‘11
- Workshop on Internet Economics
  - Sep ’09, Dec ‘11
- Joint workshop with WIDE (Japan) / CASFI (Korea)
  - Aug ’08, Apr ’09, Apr ’10, Dec ’11
- Darkspace and UnSolicited Traffic Analysis (DUST)
  - May ’12
- Cyber-security Research Ethics Dialog & Strategy (CREDS)
  - May ’23

CAIDA's blog

http://blog.caida.org

- *IPv6: What could be (but isn't yet)* Matthew Luckie
- *Shutting the phone network off while you're running out of internet protocol numbers* kc claffy
- *2001:deb:7ab1:e::effe:c75* Rob Beverly
- *Syria disappears from the Internet* Alistair King
- *Packet Loss Metrics from Darknet Traffic* Karyn Benson
- *Internet Censorship Revealed Through the Haze of Malware Pollution* Josh Polterock
- *Correlation between country governance regimes and Internet address allocations* Bradley Huffaker
- *Twelve Years in the Evolution of the Internet Ecosystem* Amogh Dhamdhere

Questions?

- publications
  http://www.caida.org/publications/papers
- collections
  http://www.caida.org/data/overview/
- workshops
  http://www.caida.org/workshops/

kc claffy
CAIDA, UC San Diego
kc@caida.org