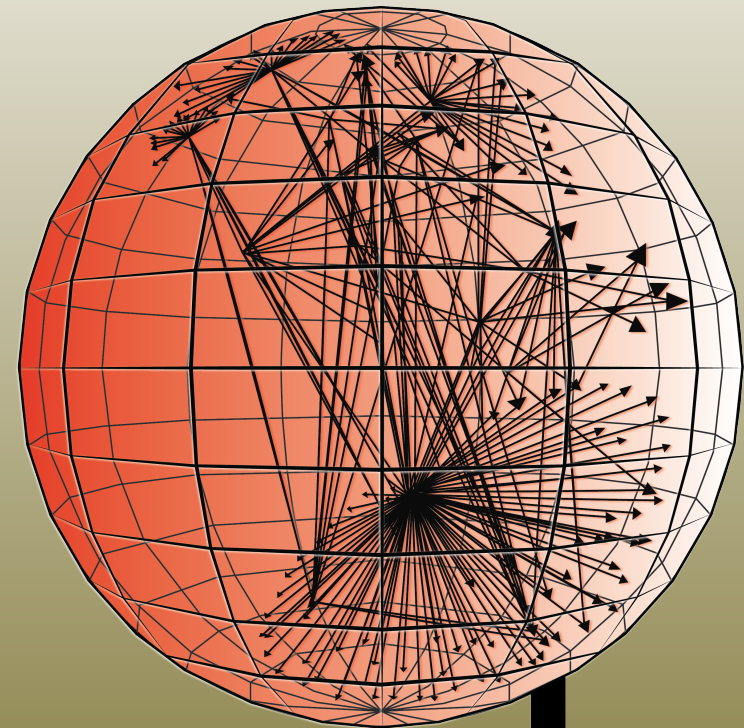*IMPACT PI Meeting*
SRI International,
Washington DC,
August 28, 2019
PI k claffy, CAIDA

TTA-1 Activities

➢ **Generate New Datasets**

➢ **Provide Data**

➢ **Host Datasets**

➢ **Contribute to Impact Team Activities**

Technical Accomplishments
Jan-Aug 2019

➢ Collecting daily scamper probes from active IPv4 and IPv6 Ark monitors, producing ~ 500 raw and post-processed topology measurement files

➢ Adding ~2TB/day of compressed Internet Background Radiation trace data (UCSD Telescope network)

➢ In January captured the last passive trace (10 GB link)

➢ About 40 IMPACT data access requests granted

➢ More than 200 Passive traces requests granted (outside of IMPACT)

➢ About 80 TB of data downloaded by all users (including 7 TB by IMPACT users)

➢ 2 New IBR Datasets indexed in IMPACT

➢ 2019-01 Internet Topology Data Kit added

➢ Updates to Interactive analysis and visualization tools shared in IMPACT (AS Rank, Vela, Henya)

Publications and Presentations
Jan-Aug 2019 (TTA-1 and TTA-2)

- ➢ 6 Publications cited in IMPACT
- ➢ 1 Presentation at [FTC Hearing](#) -- March
- ➢ 1 Presentation at the [QUILT](#) meeting – Feb.
- ➢ 4 Presentations at [DHS 2019 S&T Cybersecurity and Innovation Showcase](#) -- March
- ➢ 6 Presentations at [AIMS](#) – April
- ➢ 1 Presentation at [Horizon 2020 CANVAS workshop](#) -- March
- ➢ 1 Presentation at [PAM](#) -- March
- ➢ 2 Presentations at [USENIX](#) -- February

Resource ROI Assessment

Providing Data

CAIDA advertises 72 datasets:

- ➤ 13 paper supplements
- ➤ 15 completed experiments
- ➤ 24 One-time snapshots
- ➤ 2 Industry samplers

23 datasets require vetting via IMPACT

- ➤ 10 restricted datasets
- ➤ 13 unrestricted datasets

3 tools are shared in IMPACT

- ➤ 2 restricted tools: Vela and Henya
- ➤ 1 unrestricted tool (AS Rank)

Internet Security

Three UCSD Network Telescope Background Radiation Datasets (ongoing since 2008)

➢ Raw Background Radiation traffic traces
➢ Aggregated Flow Tuple dataset -- contains most important header fields, easy to analyze
➢ Daily RSDoS Attack Metadata

Usage:
➢ Outages
➢ Scanners
➢ Malware, e.g. Mirai Botnet
➢ RSDoS attacks

Competitors: None. No Other Project like this

Internet Traffic

Anonymized Passive Traces from 10GB Links
(caida.org/data/passive/passive_dataset.xml)

- Equinix San Jose (2008 – 2014)
- Equinix Chicago (2008 – 2016)
- Equinix NY (March 2018 – January 2019)

Usage:

- Traffic modeling
- Prototyping 100 GbE FPGA flow exporter
- Anomaly Detection and Mitigation
- Testing of security technologies

Competitors: None. Nowhere else is such data available

Internet Topology

ARK Platform: Topology Measurements (http://www.caida.org/projects/ark/)

- 190 nodes in 146 ASes
- 141 cities – 56 countries
- 78 IPv6 enabled

Usage:

- Router-level mapping
- Spoofing
- Interdomain congestions data

Competitors: RIPE Atlas. Only provides topology data. None of the above usage possible

Traceroute Queries

Henya and Vela – Systems for querying and visualizing massive archives of traceroute data

Usage:

➢ Select traceroute paths containing specified targets (e.g. IP addresses/prefixes, AS numbers, countries)

➢ Various queries, e.g. all IP prefixes announced by a given AS in BGP

➢ Geolocate to specific country

➢ RTT measurements, e.g. Interactive annotated visualization

Competitors: None

Henya and Vela Use Case
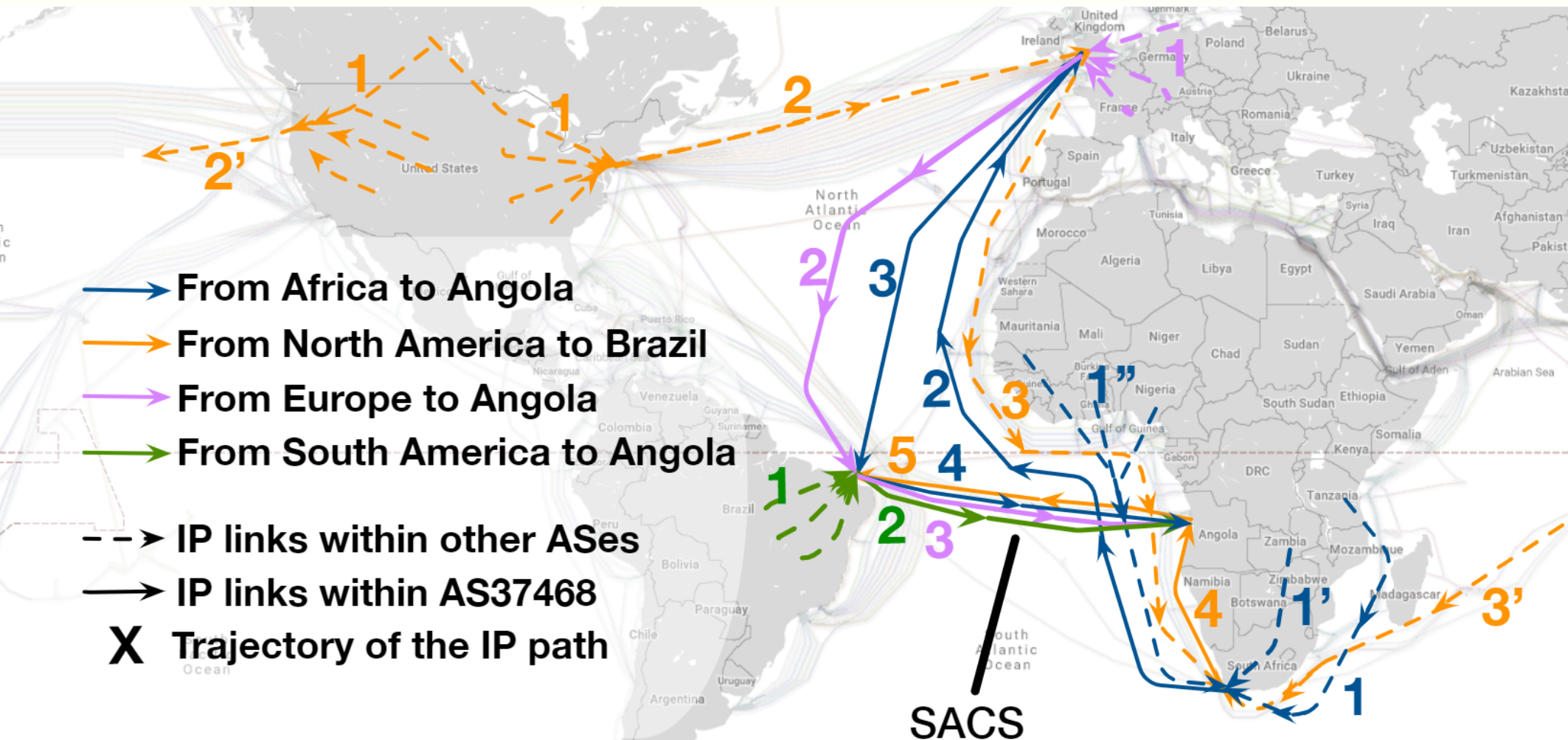
SACS deployment between Africa and S. America. Comparison between suboptimal historic trajectories followed by most IP paths and straightforward trajectory generated during on-demand run from an Ark probe after SACS deployment.

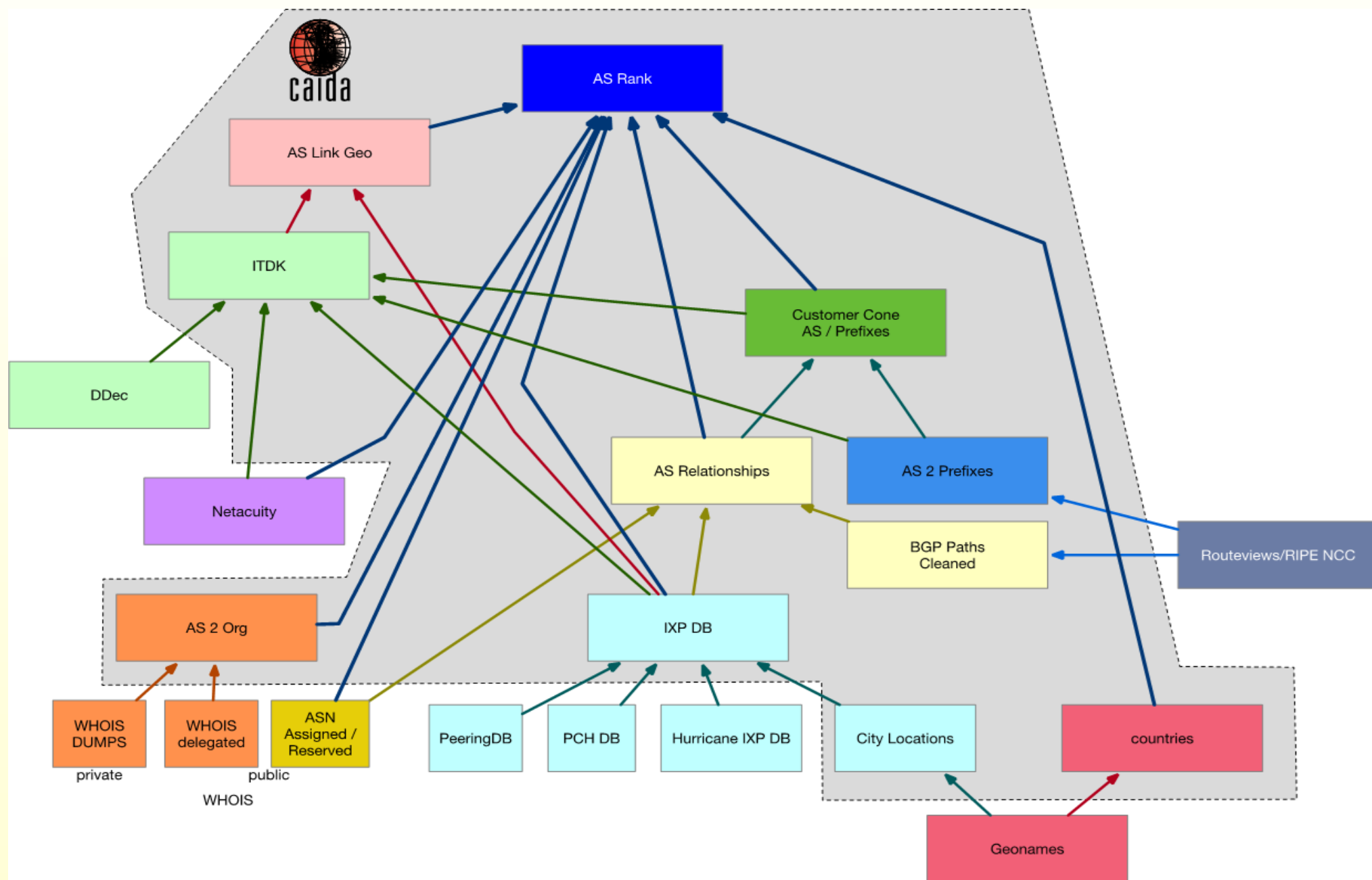# AS Rank Tool

AS Rank – BGP Data Analysis System

To explore routing and business relationships between Internet Services Providers (ASes) and organizations that own them

Usage:

- ➢ Study network robustness
- ➢ Traffic engineering
- ➢ Measurement strategies
- ➢ Economic modeling of topology

Competitors: None

AS Rank Data Flow Architecture

Success Measures

Testimonials and Letters of Support

"CAIDA tools address long-standing gaps in our ability to use large-scale data archives of Internet path data"

"CAIDA data helps improve our research on Internet topology discovery "

"We integrated CAIDA data into our experiments on congesting probing "
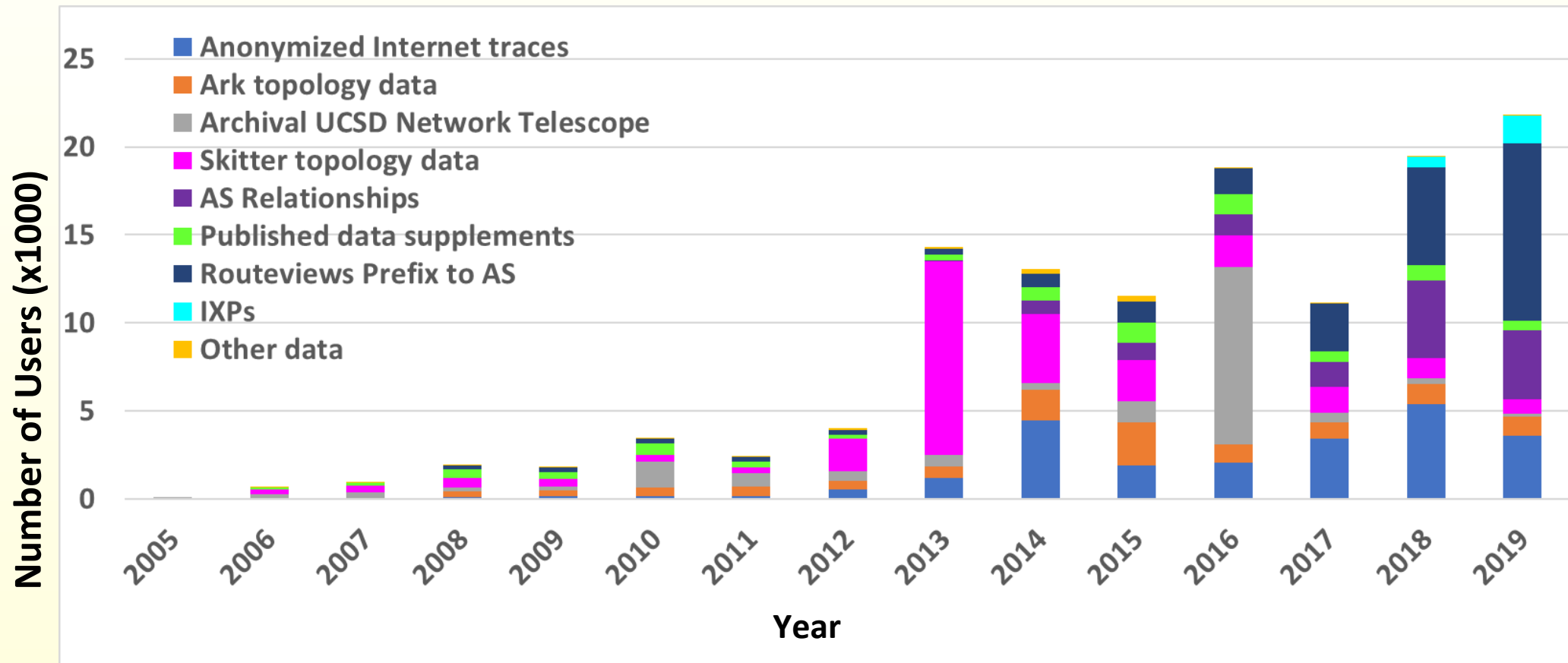
"CAIDA data enhance our ability to do Internet interdomain routing research"

"CAIDA data enables our analysis of massive archives of IP level paths"
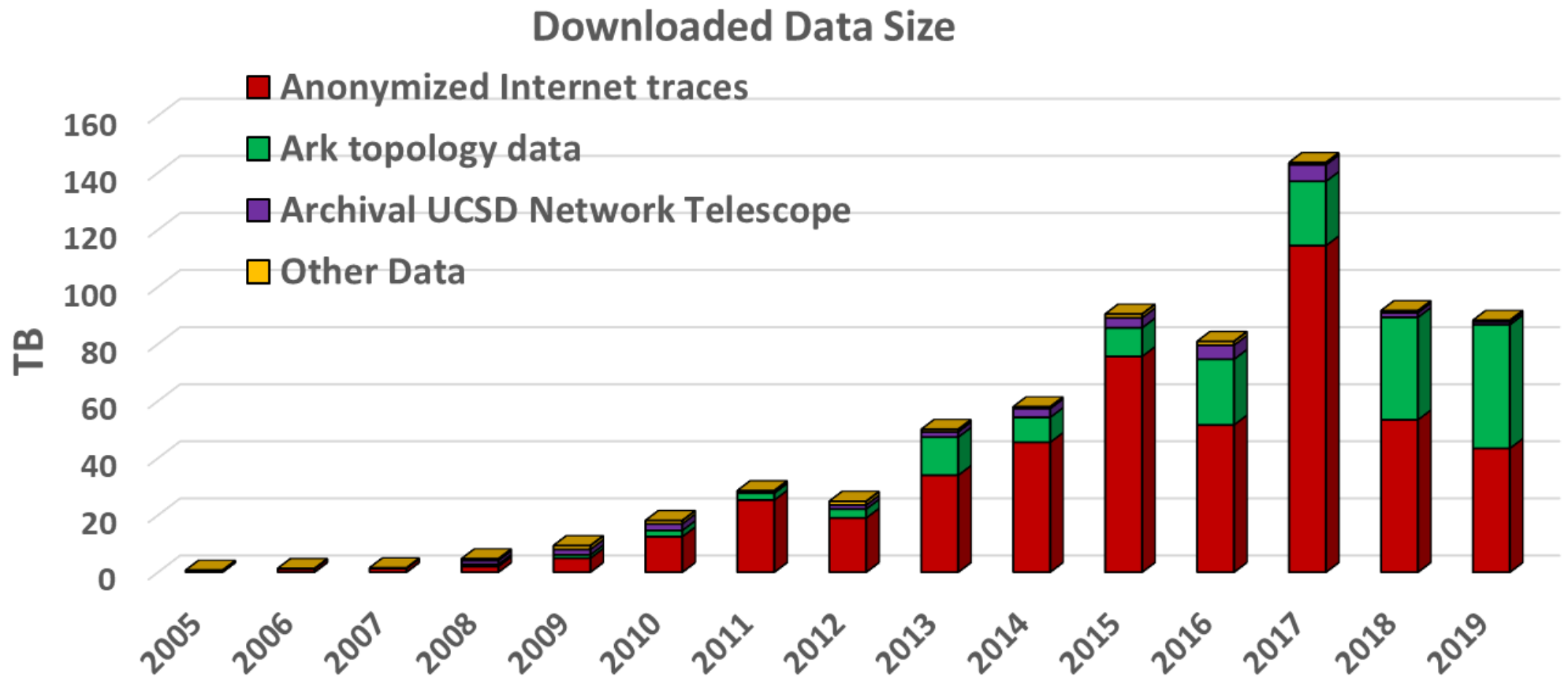
Number of Users

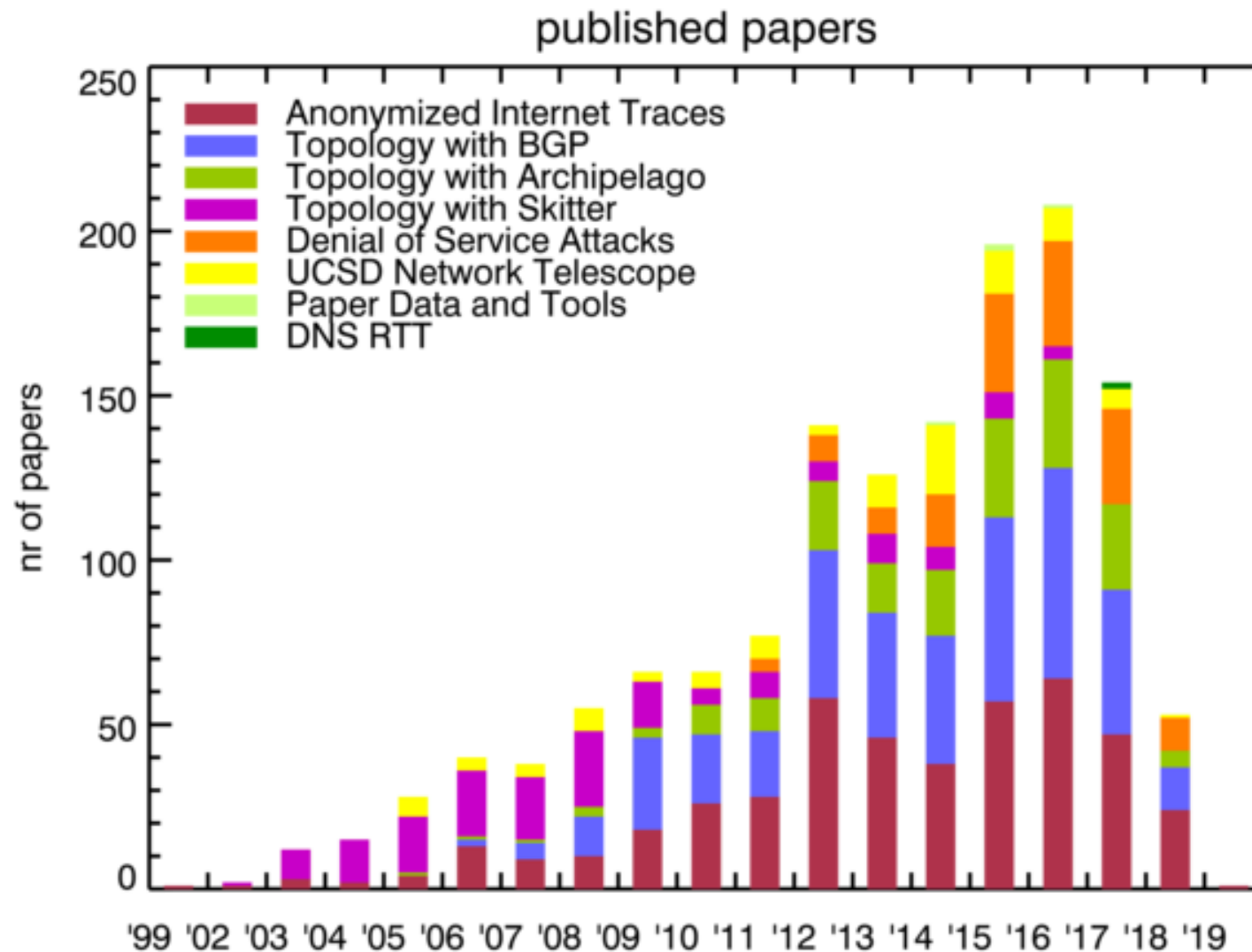## Number of unique Users (x1000)

Number of unique users downloading CAIDA data. More than 20,000 users downloaded CAIDA data in Jan-Aug 2019

Volume of Downloaded Data

Volume of data downloaded annually (Jan-Aug for 2019). Colors indicate different datasets. Multiple downloads of the same file by the same user, which is common, only counted once.

Numbers of Publications Using CAIDA Data

published papers

Legend:
- Anonymized Internet Traces
- Topology with BGP
- Topology with Archipelago
- Topology with Skitter
- Denial of Service Attacks
- UCSD Network Telescope
- Paper Data and Tools
- DNS RTT

y-axis: nr of papers (0, 50, 100, 150, 200, 250)

x-axis: '99 '02 '03 '04 '05 '06 '07 '08 '09 '10 '11 '12 '13 '14 '15 '16 '17 '18 '19

Between 2002 and July 2019 more than 1500 non-CAIDA papers using CAIDA datasets were published. These publications were cited more than 30,000 times, including about 600 mentions in various patents.
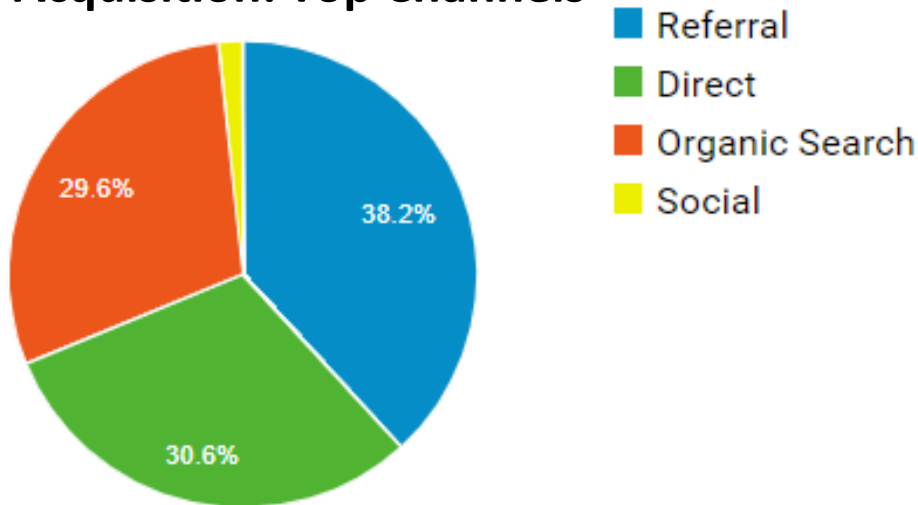
Increasing Number of IMPACT Requests
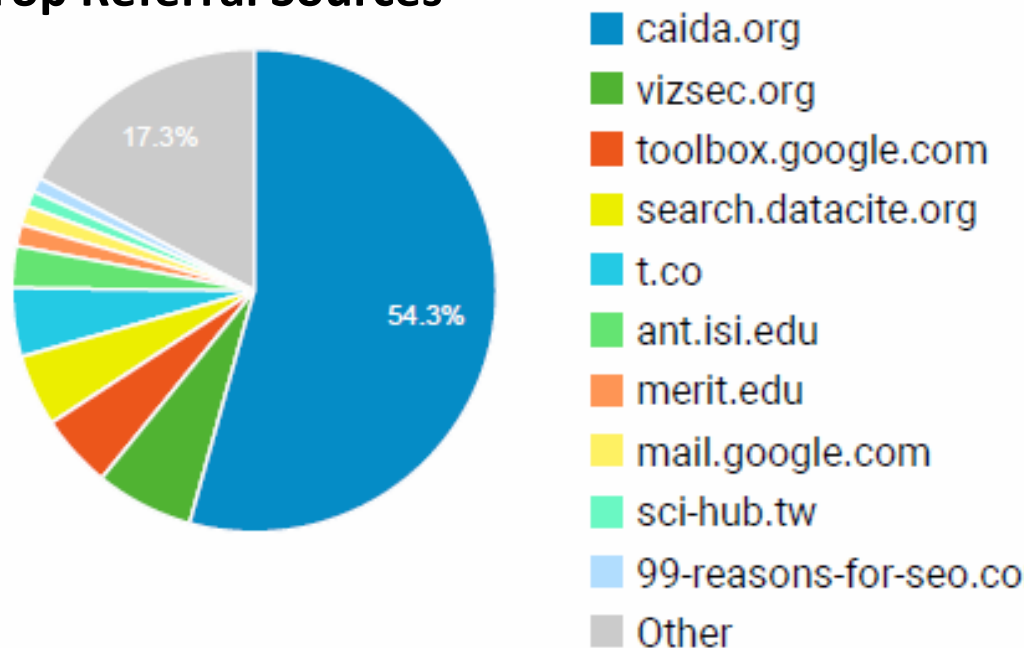
IMPACT REQUESTS
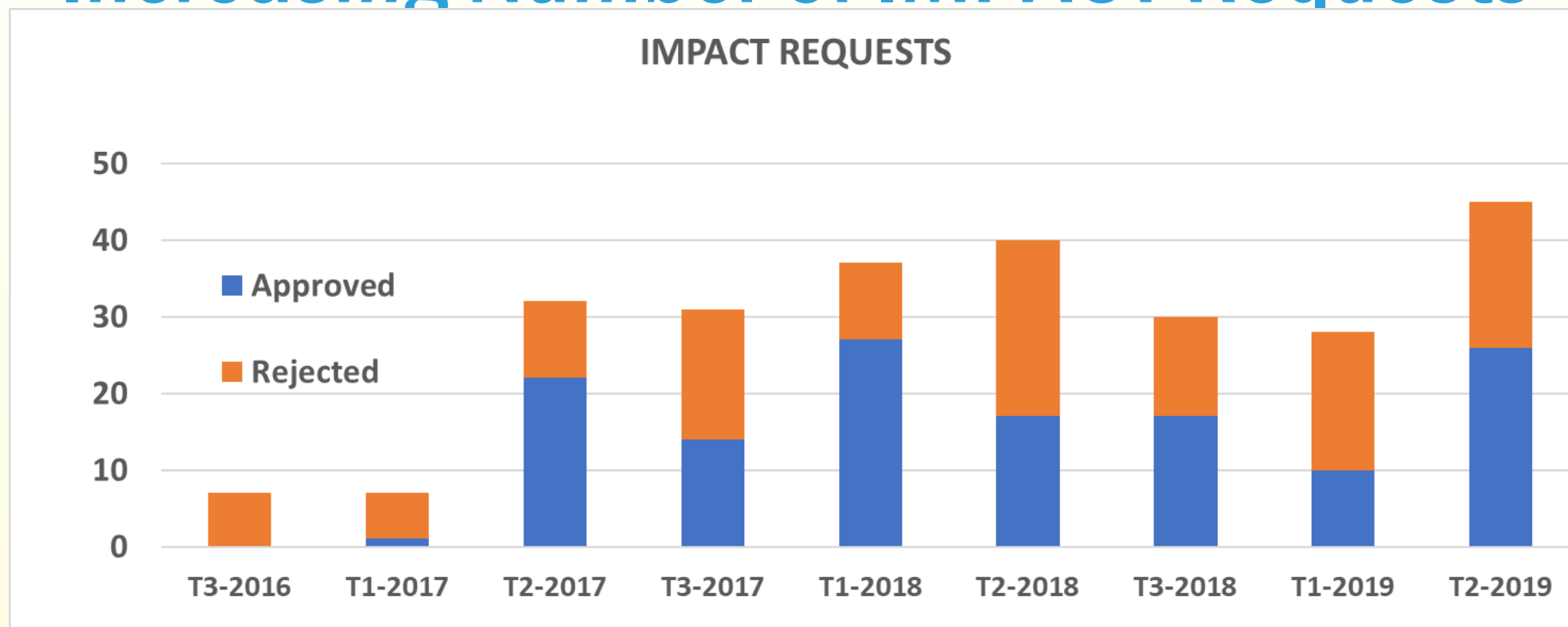
The most common reason for rejection is "Executed MOA (access to restricted datasets) has not been returned". Since 57% of 2017-2018 requests for restricted data were for CAIDA datasets, CAIDA is experiencing high rate of rejections

Attracting Visitors to IMPACT

Acquisition: Top Channels

- Referral — 38.2%
- Direct — 30.6%
- Organic Search — 29.6%
- Social

Top Referral Sources

- caida.org — 54.3%
- vizsec.org
- toolbox.google.com
- search.datacite.org
- t.co
- ant.isi.edu
- merit.edu
- mail.google.com
- sci-hub.tw
- 99-reasons-for-seo.co
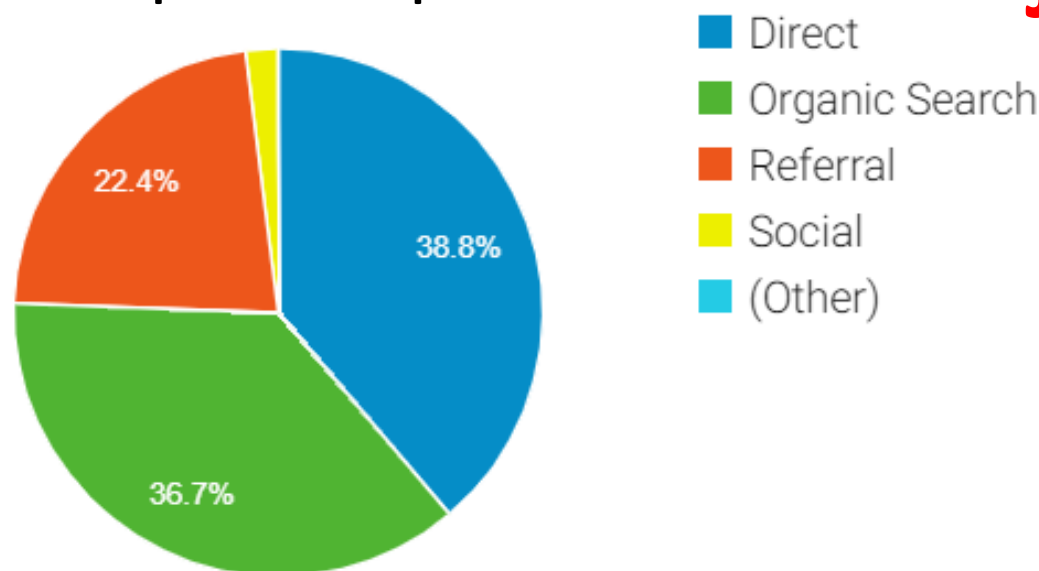- Other — 17.3%

January – December 2018

More than 20% of connections to the IMPACT website are referrals from CAIDA.org

Source:
IMPACT portal Google Analytics

Attracting Visitors to IMPACT
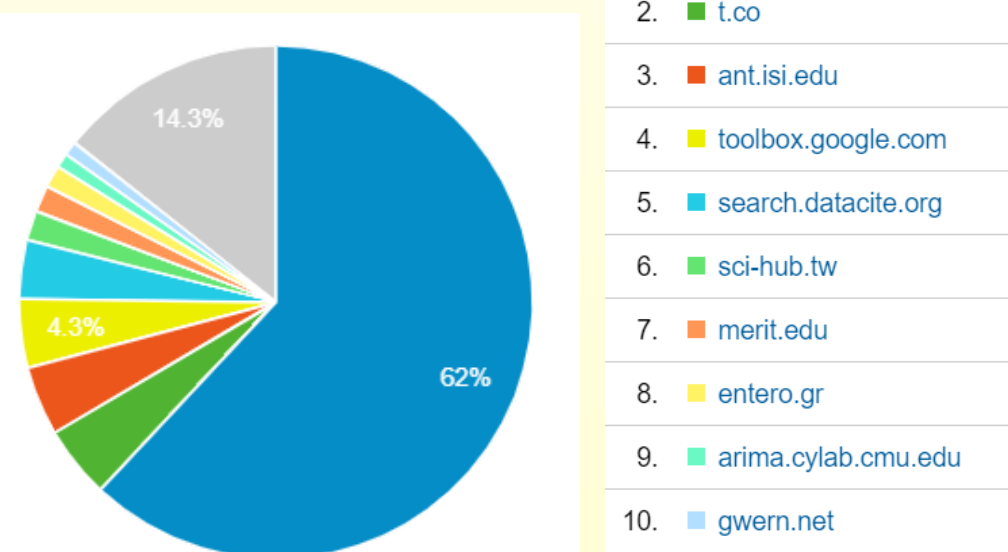
Acquisition: Top Channels

- Direct — 38.8%
- Organic Search — 36.7%
- Referral — 22.4%
- Social
- (Other)

January – August 2019

Compared to 2018:
~10 % Decrease in referral traffic
~10% Increase in organic search

More than 60% of referral traffic to the IMPACT website come from CAIDA.org

Source:
IMPACT portal Google Analytics

Top Referral Sources

1. caida.org — 62%
2. t.co
3. ant.isi.edu
4. toolbox.google.com
5. search.datacite.org
6. sci-hub.tw
7. merit.edu
8. entero.gr
9. arima.cylab.cmu.edu
10. gwern.net

(14.3%, 4.3%)

Forthcoming Developments

➢ Continue Ark and Telescope data collection

➢ Collect Passive traces at 100 GB Link

➢ Provide access to NRT raw pcap telescope data via SWIFT

➢ Produce new datasets (e.g. ITDK)

➢ Improve Interactive tools, e.g. incorporate manual annotation into AS Rank tool (Operator feedback, FCC filings, reports, etc)
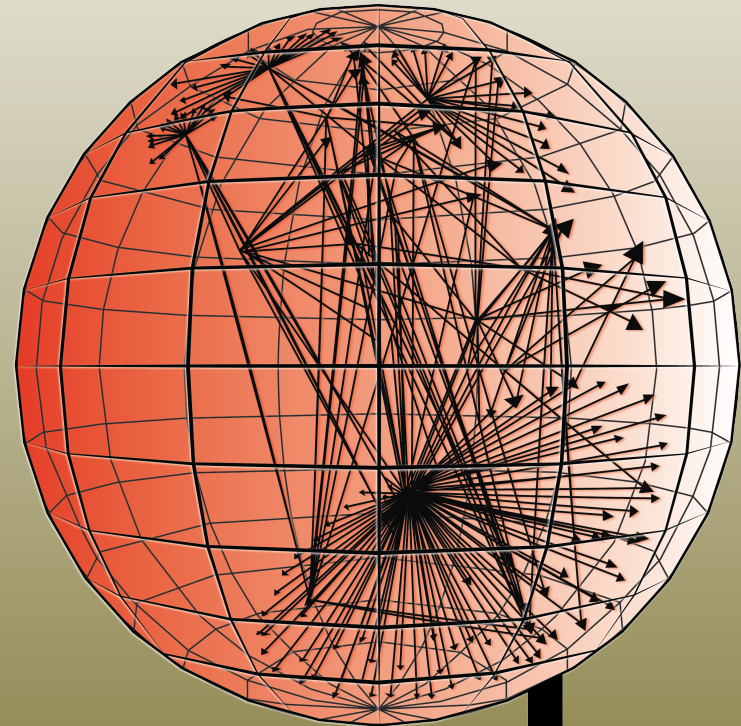
Contact Information

PI: k claffy, CAIDA
kc@caida.org
http://www.caida.org/