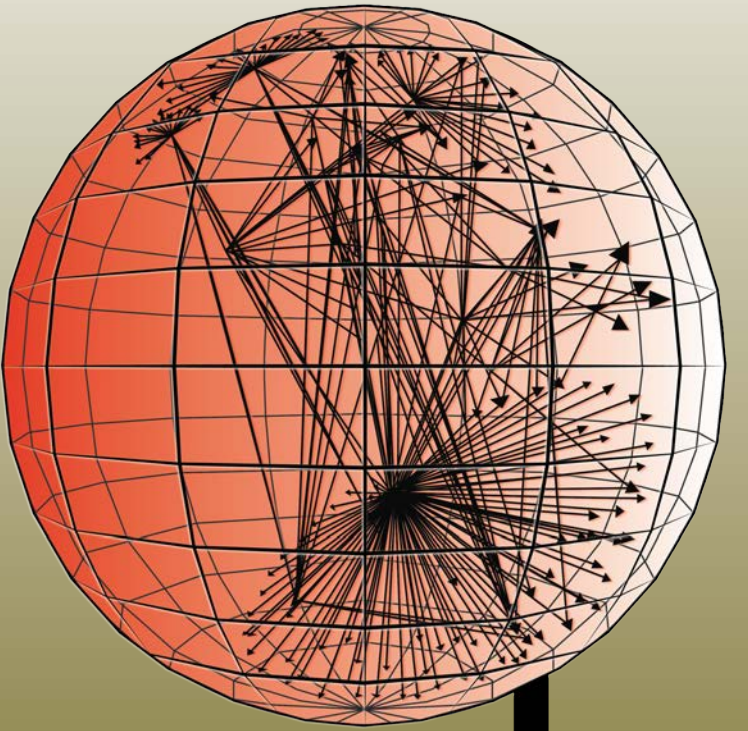
IMPACT PI Meeting
Virtual,
March 23, 2020
TTA1 PI k claffy, CAIDA



caida

TTA-1 Activities

- Generate New Datasets
- Provide Data
- Host Datasets
- Contribute to Impact Team Activities

Technical Accomplishments
September 2019 – March 2020

- Collecting daily scamper probes from active IPv4 and IPv6 Ark monitors, producing ~ 700 raw and post-processed topology measurement files
- Adding ~2TB/day of compressed Internet Background Radiation trace data (UCSD Telescope network)
- About 30 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)
- 2019-04 and 2020-01 Internet Topology Data Kit added
- Developed Dory.caida.org -- Interactive portal for Ark monitors management (supported by GraphQL)

Publications and Presentations

Sept 2019-March 2020 (TTA-1 and TTA-2)

- 8 Publications cited in IMPACT
- 1 Presentation at NIST -- January
- 2 Presentations at the IMC meeting -- Oct
- 2 Presentations at DUST -- Sept
- 2 Presentations at AIMS/KISMET -- Feb
- 2 Presentations at WIE-KISMET -- Dec
- 1 Presentation for IIJ (Japan) -- Jan
- 1 Presentation for M^{3}AAWG -- Jan

Providing Data (73 datasets 23 in IMPACT)

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

performance

DNS root/gTLD RTT DATA

security

Code Red Worms, Backscatter, DDoS attacks, Witty Worm, Conflicker

topology

AS Links, Prefix to AS, AS Rank, AS Relationships, Archipelago IPv4+IPv6 topology

topology (processed)

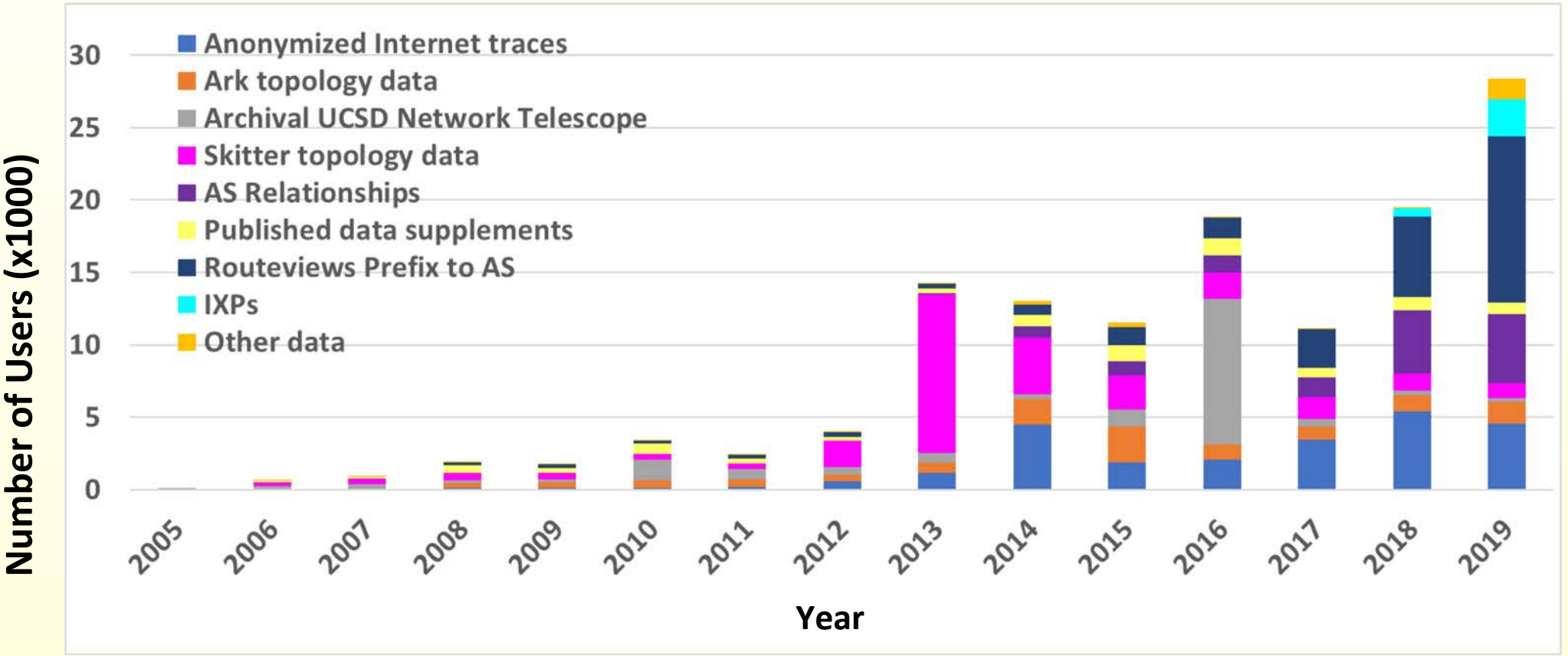
Macroscopic Internet Topology Data Kit (ITDK)

traffic

Telescope Data, Telescope (live), Anonymized Internet Traces, Tier 1 packet traces, SDNAP

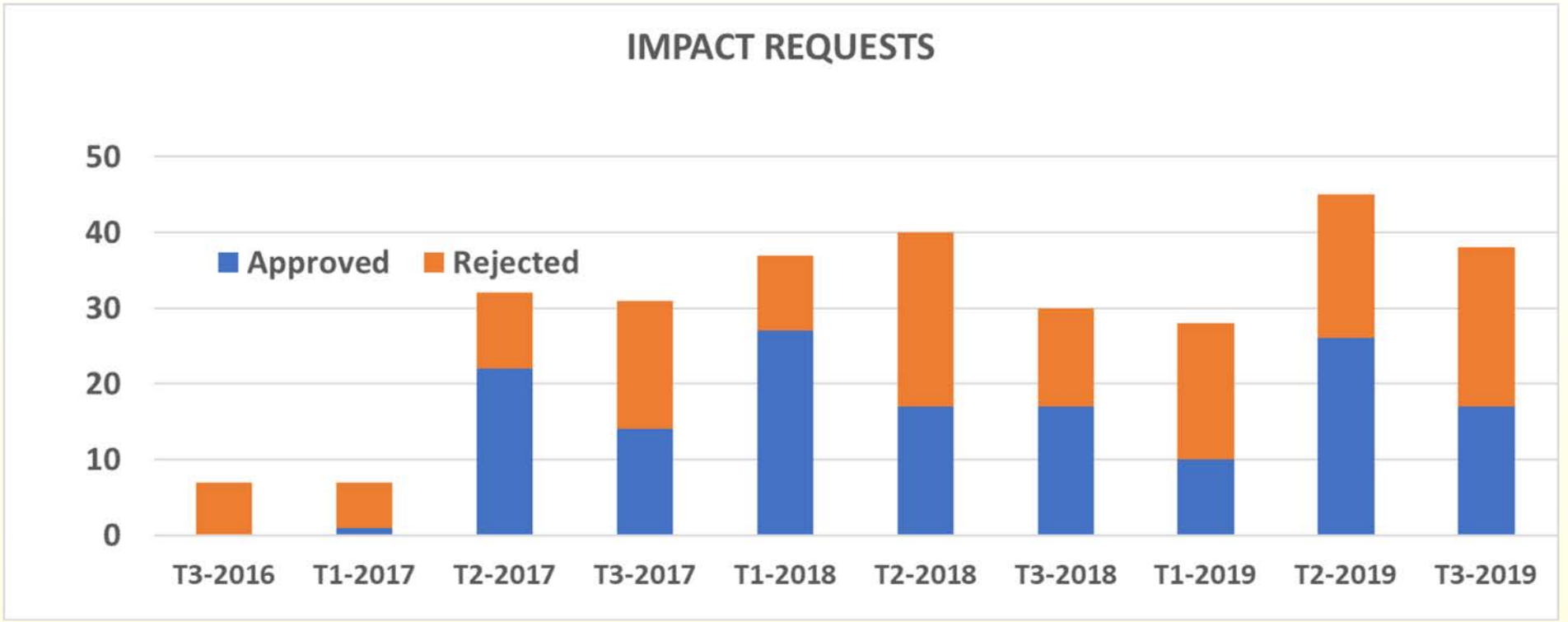
Number of Users

Number of unique Users (x1000)



Number of unique users downloading CAIDA data. More than 25,000 users (unique Ips) downloaded CAIDA data in 2019

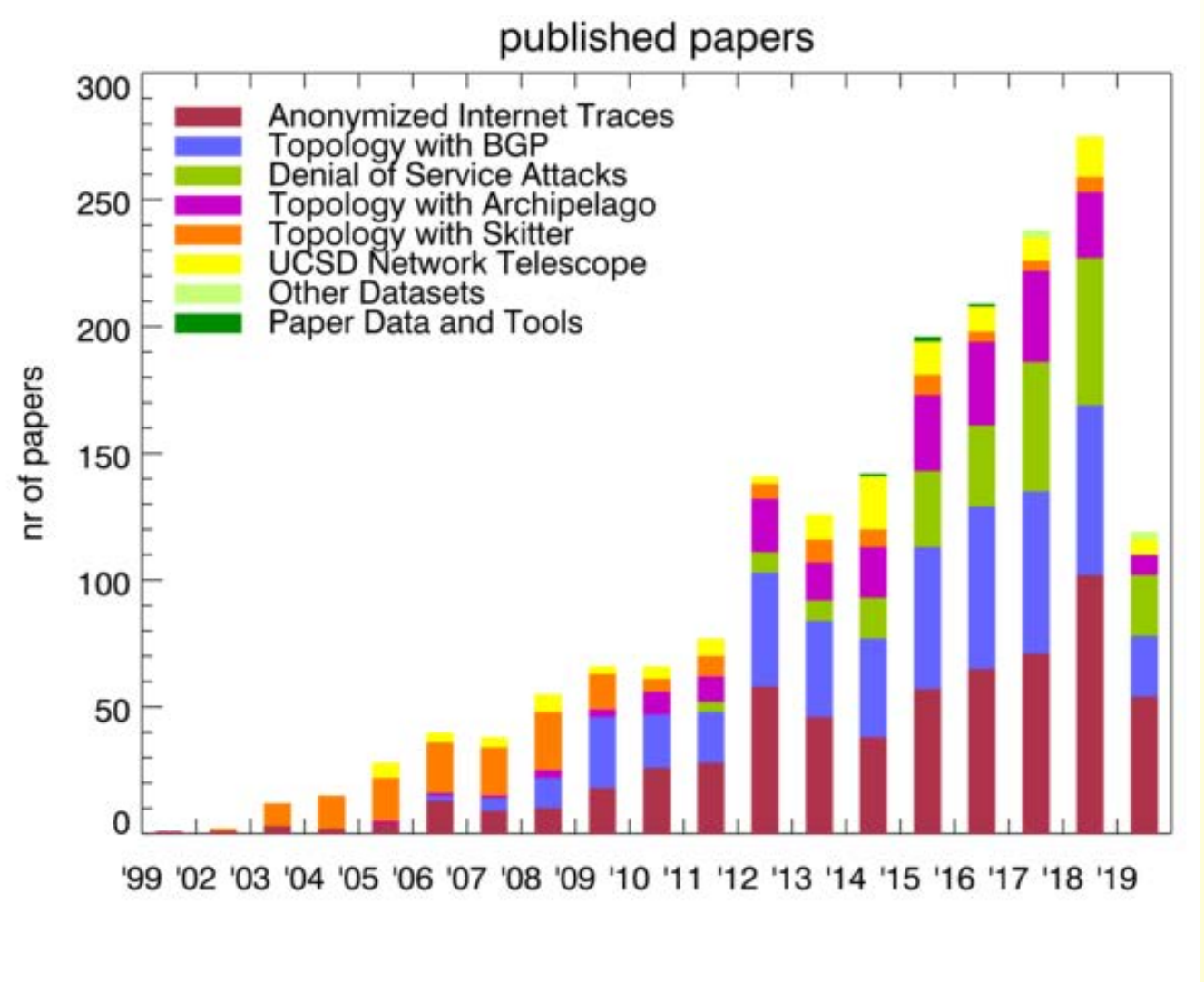
Number of IMPACT Requests



Most common reason for rejection: "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. *(Should not be called "rejected"..)*

Numbers of Publications Using CAIDA Data (as of Sept-2019)



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.

Internet Security

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

https://www.caida.org/data/passive/telescope-near-real-time_dataset.xml

- Raw Background Radiation traffic traces
- Aggregated Flow Tuple dataset -- contains most important header fields, easy to analyze
- Daily RSDoS Attack Metadata (2008 – June 2019)

Usage:

Outages; Scanners; Malware, e.g. Mirai Botnet; RSDoS attacks

Competitors: None. No Other Project like this

Internet Security

UCSD Telescope datasets

- Adding ~ 2 TB/day
- Analyzed on CAIDA machines – have started to create VM for each user
- Flow Tuple and Daily RSDoS -- all data stored in SWIFT -- OpenStack Object Storage system at CAIDA/UCSD
- Raw pcap data -- last 30+ days are in SWIFT

Internet Traffic

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

- The only restricted dataset accessed directly through CAIDA
- Last trace collected in January 2019
- More than 300 new users in 2019
- More than 60 Tb downloaded in 2019
- Plan -- Move to SWIFT, provide access via Globus

Usage:

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

Competitors: None. Nowhere else is such data available

Internet Traffic

Acquire and Deploy 100GB packet capture monitor

Measurement machine:

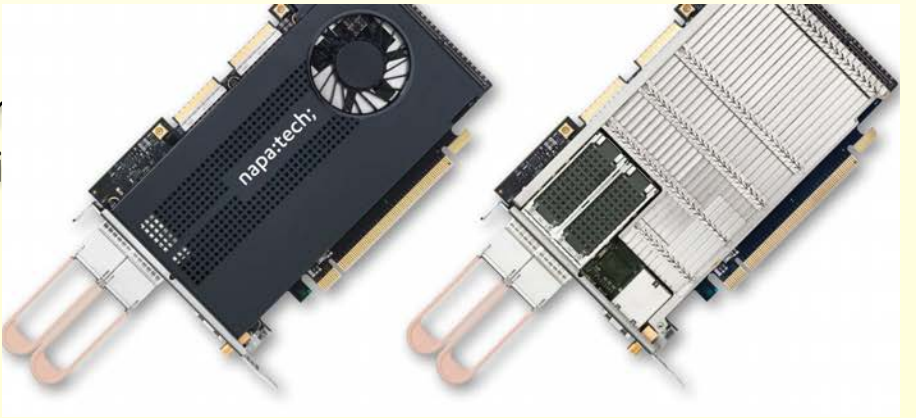
- Dual AMD EPYC ROME 7352, 24-core, 2.3GHz Processors; 256GB RAM;
- Two 480GB SATA SSDs for OS; Mellanox MCX516A-CCAT Dual Port 100GbE QSFP28 Network Adapter for streaming to Analysis machine

Analysis machine:

- (2) EPYC Rome 7452 2.35GHz Thirty-Two Core; 256GB RAM; (8) 1.6TB NVMe SSDs, 3DWPD Endurance for capturing data; (2) 240GB SATA (RAID1) for OS;
- Two 1/10GBase-T Ports (onboard); (1) Dual Port 100GbE NIC (MCX516A-CCAT) for accepting stream from analysis machine

Capture card:

- NAPA:tech Link™ NT200A02 SmartNIC car
- Mellanox MCX516A-CCAT card – alternati



Backbone Link:

- DREN 100 GB

Internet Topology

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

- 177 nodes in 134 ASes
- 137 cities – 54 countries
- 70 IPv6 enabled

Usage:

Router-level mapping

Spoofing

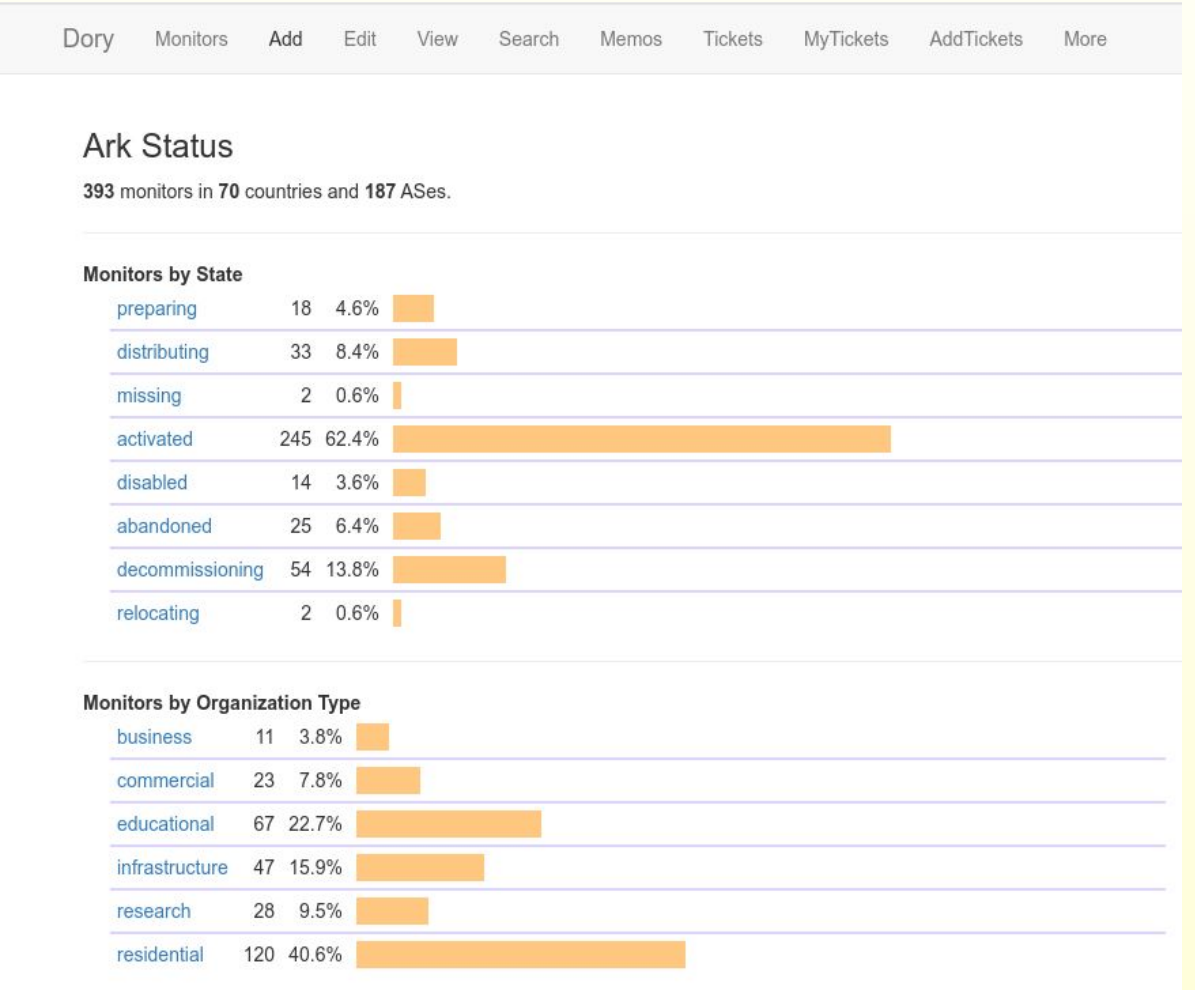
Interdomain congestions data

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

Ark Monitors Management

DORY – dory.caida.org

- GraphQL API
- GraphQL client in Python
- Adding new Monitors
- Updating info for existing monitors
- Creating tickets and memos



Ark Topology data

2019 Metrics

- Added about 150K new files
- 17 New IMPACT users
- About 10 TB Downloaded by IMPACT users
- Added 2019-01, 2019-04 and 2020-01 ITDK (https://www.caida.org/data/internet-topology-data-kit/) datasets

Traceroute Queries

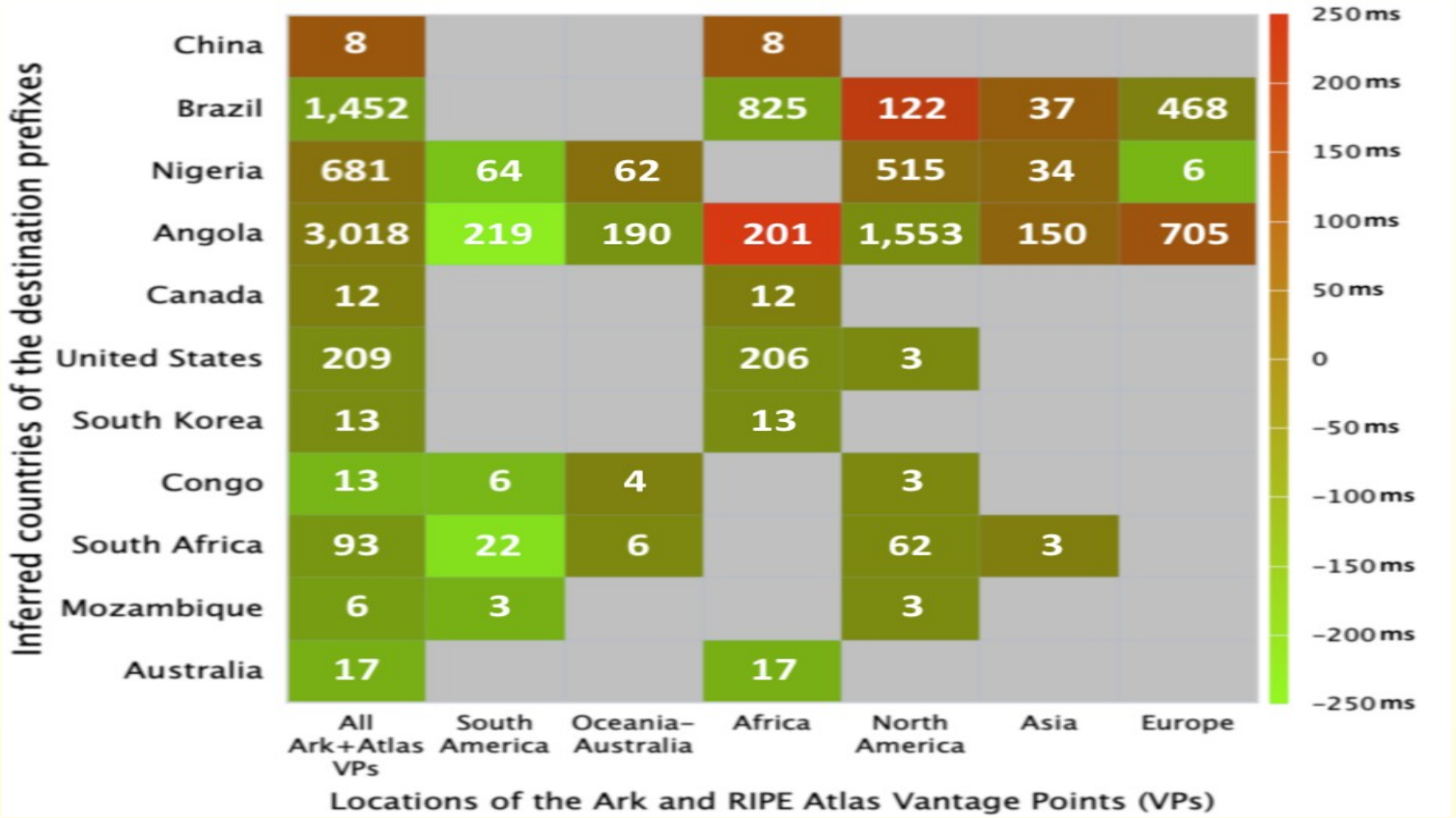
Henya and Vela – Systems for querying and visualizing massive archives of traceroute data

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. Change in RTT (After - Before) of the medians of minimum RTTs per week pre&post SACS for observed < **s**ource; **d**estination > pairs. Each cell contains the number observed < s; d > pairs, colored according to the corresponding RTT; The highest performance improvements -- from South America to Angola or South Africa, the worst degradations -- from Africa to Angola or North-America to Brazil.*

Forthcoming Developments

- Continue Ark and Telescope data collection
- Collect Passive traces at 100 GB Link
- Move passive traces data to SWIFT
- Provide access to data in SWIFT via Globus
- Produce new datasets (e.g. ITDK)
- **Create an open Knowledge of Internet Structure: Measurement, Epistemology, and Technology** (KISMET) network

Open Knowledge Network

National Science and Technology Council Open Knowledge Network Vision

October 4-5, 2017 Big Data IWG workshop

"A knowledge network allows stored data (both structured and unstructured data) to be located and its attributes and relationship to other data and to real-world objects and concepts to be understood at a semantic level. Today, technology companies develop largely proprietary knowledge networks, often specialized for customer needs. Instead, this Open Knowledge Network will build public-private cooperation and engage convergence teams from all areas of data science and science and engineering domains to create a shared, open infrastructure." **— October 4-5, 2017 NSF Convergence Accelerator Pilot (NSF C-Accel)**

CAIDA proposed to: Develop an Open Knowledge Network (OKN) of public data on Internet structure, i.e., the naming, addressing, and routing systems, to confront a growing empirical gap in science, security, and public communications policy.

OKN-KISMET at a Glance

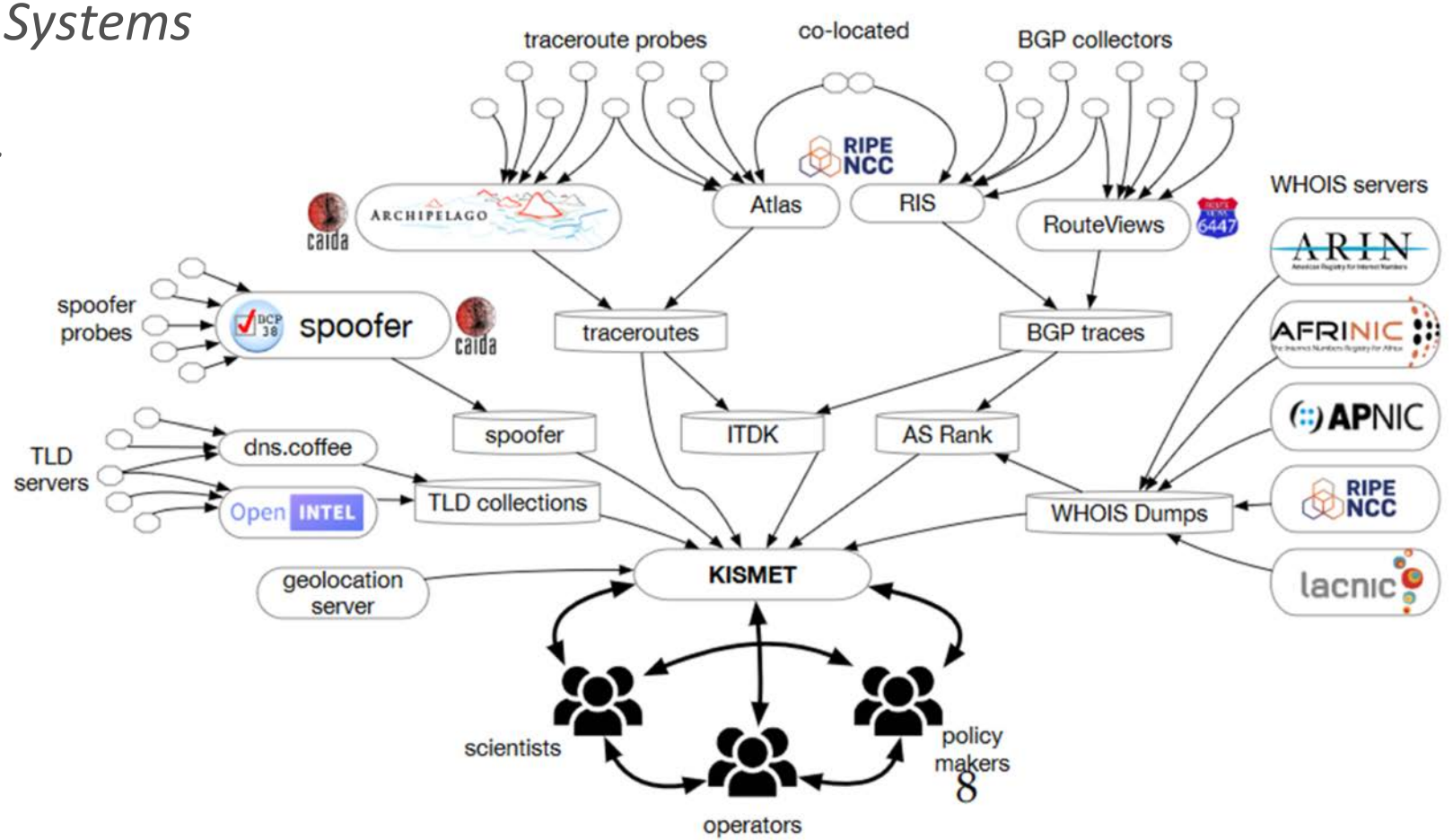
Multi-stakeholder team building effort

- *academic, government, industry*

***Initial* focus on Internet identifier systems**

Explore rich relationships across:

- *domain names*
- *Autonomous Systems*
- *IP address*
- *name servers*



OKN KISMET WORKSHOPS

- **Dec 9-11 2019: Workshop on Internet Economics – KISMET**
- **Feb 26-28 2020: Workshop on AIMS – KISMET**

Established common ground:

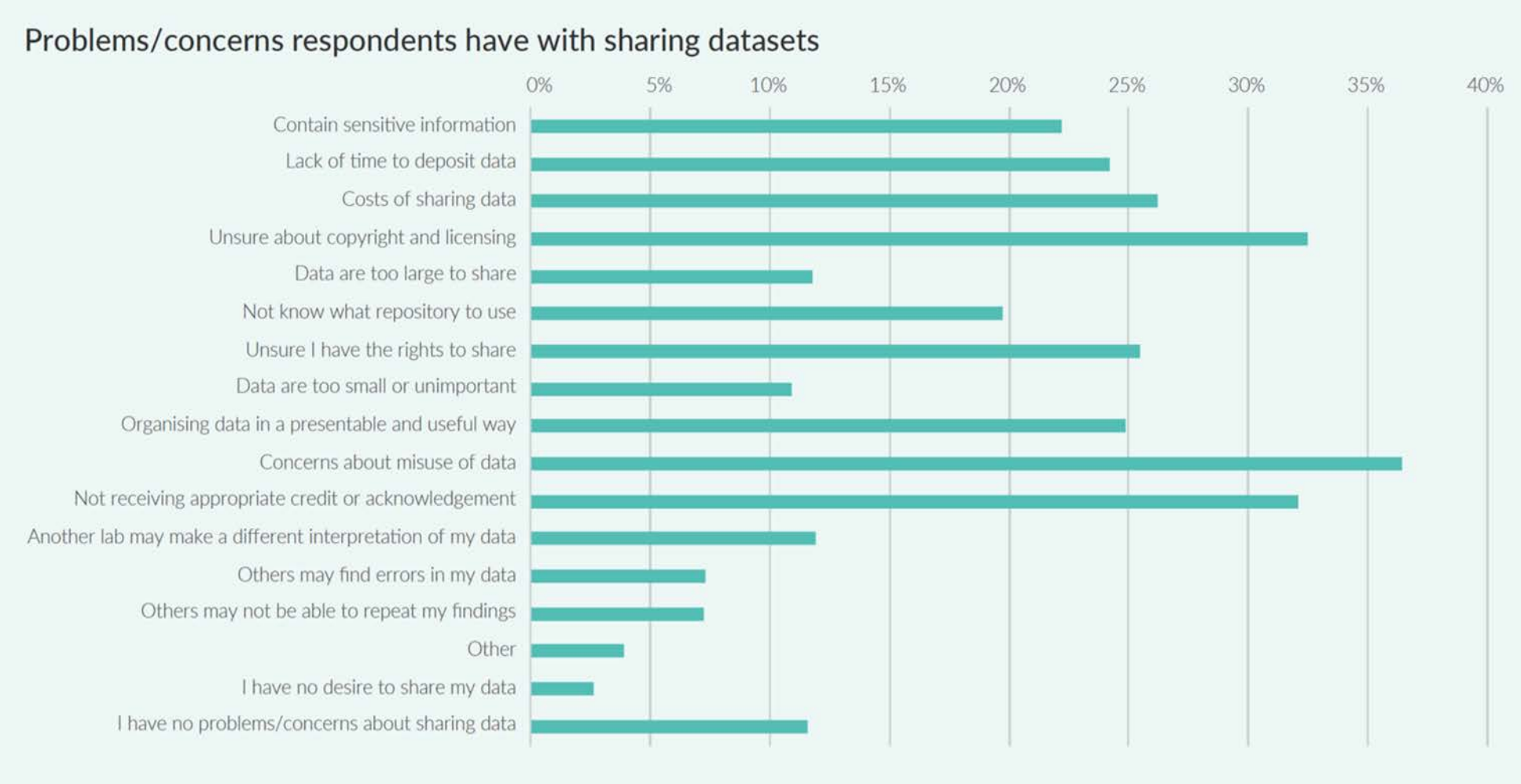
- Data availability
- Data usage and support
- Understanding of the chasms between raw data, appropriately curated data, and scientific knowledge related to Internet infrastructure security and stability

Identified Needs

- FAIR principles for Internet data management and stewardship
- Standards
- Meta-data that enables discoveries
- Knowledge Graphs
- Ontologies
- Interoperability
- Research on new Internet protocols (e.g. DOH/ABCD, M2M)
- Research of new models of Internet usage (smart cities, fog computing, 5G)

OKN Challenges

https://digitalscience.figshare.com/articles/The_State_of_Open_Data_Report_2019/9980783



Future of IMPACT

Plan of Action

- Figure out what data is needed
- Identify federal strategic priorities
- Transparency
- Document Failures and successes
- Leverage other efforts and vision
- Tech transfer from Industry to Academia
- Reproducibility (paper supplements)
- Metrics of success
- Support public interest in policy evolution

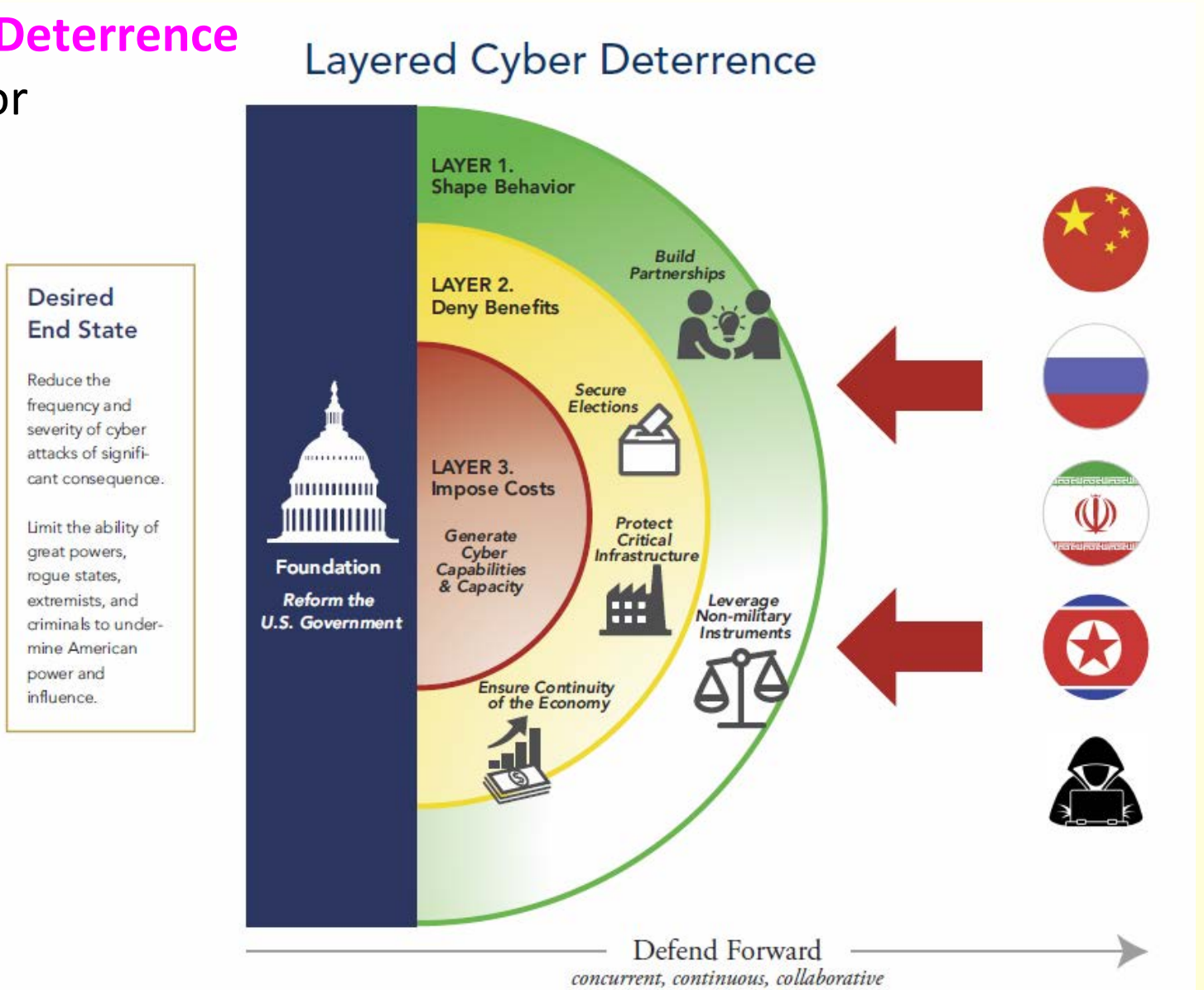
US Cyberspace Solarium report

(Example strategic priorities March 2020)

https://www.solarium.gov/home

3 Layers of Cyber Deterrence

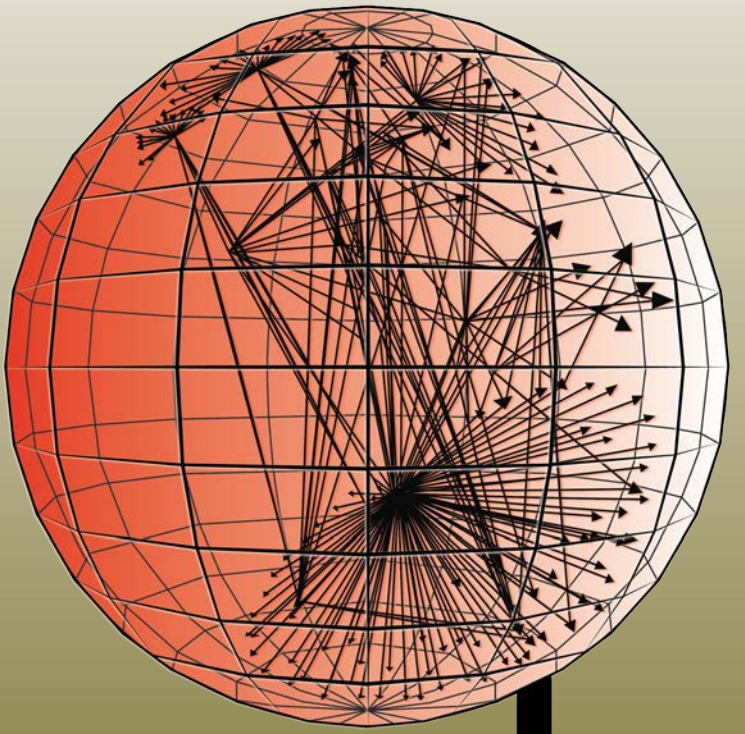
I. Shape behavior
II. Deny benefits
III. Impose cost



Recommendations

Organized in 6 Pillars:

1) *Reform the U.S. Government's Structure and Organization for Cyberspace.*
2) *Strengthen Norms and Non-Military Tools.*
3) *Promote National Resilience.*
4) *Reshape the Cyber Ecosystem.*
5) *Operationalize Cybersecurity Collaboration with the Private Sector.*
6) *Preserve and Employ the Military Instrument of National Power.*



caida

Contact Information

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