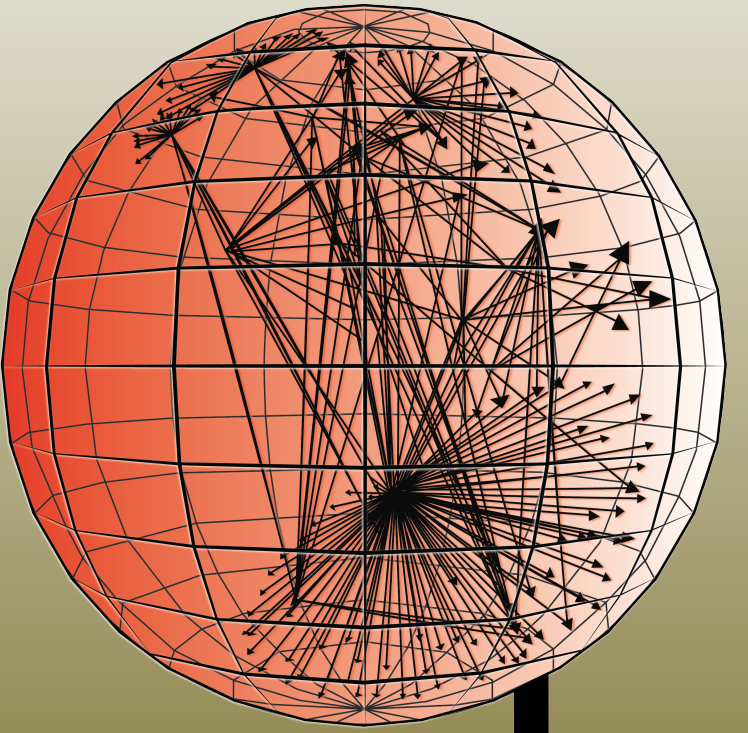
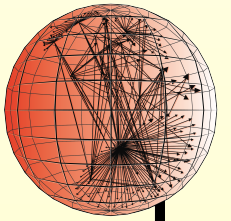
Towards an Open Knowledge Network about Properties of the Internet Identifier Systems

PI k claffy, CAIDA



caida

The Center for Applied Internet Data Analysis (CAIDA) Overview



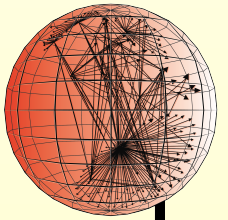
Overview

Main Activities:

- *Network research*
- *Infrastructure for large-scale Internet measurement*
- *Data curation and distribution*

Main Projects:

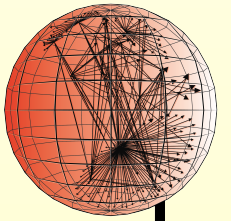
- *Mapping the Internet*
- *Mapping Interconnection Connectivity and Congestion*
- *Monitoring Global Internet Security and Stability.*
- *Future Internet Architectures*
- *Public Policy*



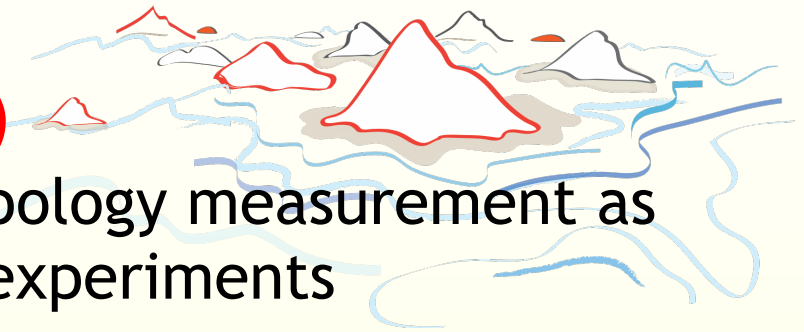
CAIDA's Mission

To Investigate Practical and Theoretical Aspects of the Internet in Order to

- Improve the integrity of the field of Internet science
- Inform science, technology, and communications public policies
- Foster a collaborative environment in which data can be acquired, analyzed, and shared
- Provide macroscopic insights into Internet infrastructure, behavior, usage, and evolution



Measurement Infrastructure



Archipelago (ark)

supports ongoing topology measurement as well as customized experiments

UCSD Internet Telescope (IBR)

packet capture to largely unused address space (one-way traffic only)

Passive Trace Capture

capture(d) packets on Tier 1 10GB backbone link (two-way traffic)
shared anonymized headers only

caida

Archipelago

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

Deployment

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

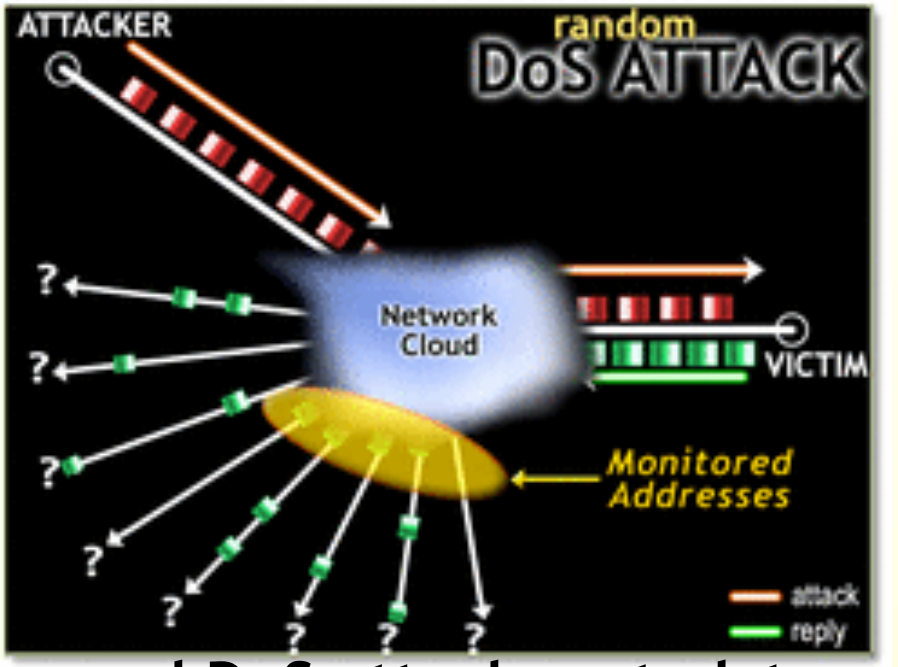
Current projects

- MANIC (89)
- Researcher experiments,
  - spoofer
  - Youtube
  - QOE experiments
- Team-probing collects IPv4 and IPv6 topology (172)

caida

UCSD Network Telescope

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



- Portion of Internet address space that is mostly unused
- 0.2% of the Internet address space
- Traffic reaching the router is *unsolicited* "Background Radiation"
- (malware,botnet,scanning,DoS,etc)

- We collect and analyze this traffic
- Daily raw pcap data, aggregated flows and DoS attack metadata
- ~17 PB compressed data stored at DOE's NERSC
- Currently adding ~ 3-4 TB/day

caida

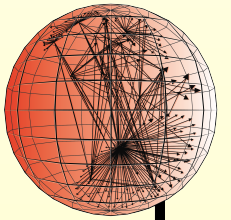
Anonymized Passive Traces from 10GB Links

http://www.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



CAIDA Services

http://www.caida.org/services

Services	Interfaces	Tags	Status
ARank	Web app / API	AS/org inter-domain relationships as-rank.caida.org	public
BGPSTREAM	API	BGP routing data analysis support bgpstream.caida.org	public
IODA INTERNET OUTAGE DETECTION AND ANALYSIS	Web app	outages, darknet ioda.caida.org	public
MANIC	Web app / API	congestion, interdomain/IP links manic.caida.org	restricted
Vela	Web app/ API	IP topology, ping, traceroute, Ark vela.caida.org	restricted
PANDA	Web app/API	Internet related database / API	under development

services

(under development)

- **Consistent, accessible interfaces to data sets**
- **Software libraries to facilitate use of data**
- **Data infrastructure building blocks to enable development of sophisticated analysis platforms**

datasets | topics | entities | joins | papers

geolocation

datasets

AS Rank topology, geolocation, ranking — 12 papers

CAIDA's ranking of Autonomous Systems (AS) (which approximately map to Internet Service Providers) and organizations (Orgs) (which are a collection of one or more …

AS names,3+ ,Organization names,3+ ,AS Link IPv4 relationship ,Country name,3+ AS+Country ,Organization+Country ,Organization+AS ,AS Link IPv4+AS, 1+

Netacuity geolocation — 35 papers

Digital Element's NetAcuity is the industry-standard for accurate, reliable and granular geolocation and IP Intelligence data.

IPv4 ,IPv6 ,City name,3+ ,IPv4+City ,IPv6+City

papers

datasets | topics | entities | joins | papers

dataset:IODA

papers

Profiling BGP Serial Hijackers: Capturing … topology, security, routing

BGP hijacks remain an acute problem in today’s Internet, with widespread consequences. While hijack detection systems are readily available, they typically rely on a priori prefix-…

Cecilia Testart, Alistair King, Alberto Dainotti, David Clark

IDOA: AS name , AS+Country

BGPStream: Prefix number bytes , AS+Prefix

solutions

datasets | topics | entities | joins | solutions | papers

geolocation

solutions

How do you find an AS's country? geolocation

The as2org files contain two different types of entries: AS numbers and organizations. The two data types are divided by lines that start with.

AS Organization: AS , Country name , AS+Country

CAIDA Data Collections (73 Datasets)

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

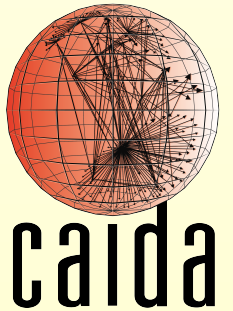
performance
- DNS root/gTLD RTT DATA

security
- Worms (Code Red,Conficker), Backscatter, DDoS attacks

topology
- AS Links, Prefix-to-AS, AS Relationships,
- IPv4+IPv6 topology (curated)
- Macroscopic Internet Topology Data Kit (ITDK)

traffic
- One-way IBR traffic (traces and live access)
- Two-way anonymized packet headers from backbone*

*Data no longer being collected; need to upgrade monitors.

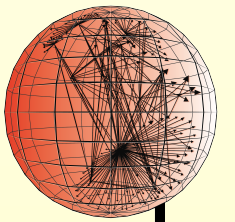


Numbers of Publications Using CAIDA Data (as of 08-01-2019)

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.

caida

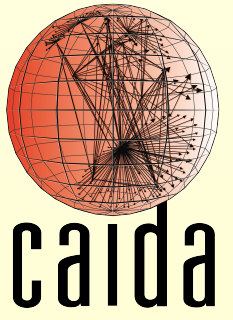
Challenges for the Future of the Internet Ecosystem



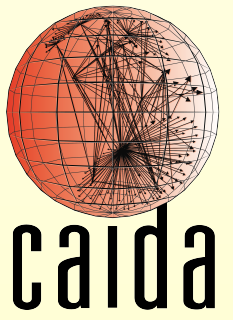
No shortage of harms

Lack of trust and privacy
Political polarization
Death of trusted journalism
Consolidation and centralization of power/capital
Slowed economic growth, productivity, innovation
National security vulnerabilities
Corruption of democratic political processes
Risk of smart toys for children
Malware
Election security
Cybercrime
Online bullying

http://www.caida.org/publications/papers/2019/toward_theory_harms_internet/



But there are also challenges under the hood of this ecosystem



Threats to Internet Infrastructure

Addressing Architecture

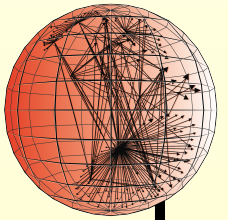
- No incentive to comply with IETF-recommended Source Address Validation
- Lack of a trustworthy registry of data on which organizations have operational authority over which IP addresses

Interdomain routing system

- Mis-announcement of BGP prefixes
- Hijacking
- Route Leaks

Domain Naming System

- Lack of authentication - DNS can generate false mappings of names to IP addresses (DNSSEC addresses but deployment incentive low)
- Name registration ecosystem support for privacy unintentionally encourages malicious actors

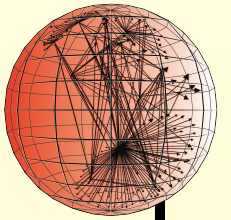


Why are these layers important?

- They underlie all activities on the Internet

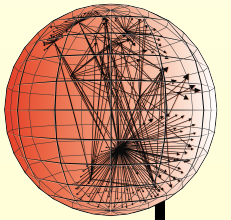
Why are threats to these layers so resistant to mitigation?

- They require some level of global governance to ensure consistent, reliable interpretation
- Current models of governance are inadequate



Barriers to mitigation

- Misaligned incentives
- Absence of Internet-ecosystem metrics
- Absence of accurate and consistent data
  (data collection is driven by operational needs,
  constrained by collection cost, hence limited use)
- No consensus on privacy vs security tradeoffs
  (e.g. GDRP,CCPA)
- Inability to articulate, identify, quantify harms
- Not clear who should regulate the Internet



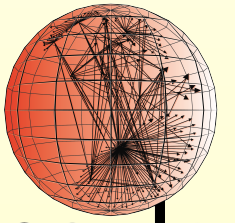
National Science and Technology Council Open Knowledge Network Vision

(to the rescue?)

"… An "open" knowledge network (OKN) would be available to all stakeholders, including the researchers who will help push this technology further. An OKN requires a nonproprietary, public-private development effort that spans the entire data science community and results in an open, shared infrastructure."

— October 4-5, 2017 Big Data IWG workshop

Is an Open Knowledge Network a potential component of a solution for Internet infrastructure threats?



Path Forward – OKN KISMET

Knowledge of Internet Structure: Measurement, Epistemology, and Technology

April 2019: NSF Convergence Accelerator Pilot (Open Knowledge Network) call for proposals

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.

caida

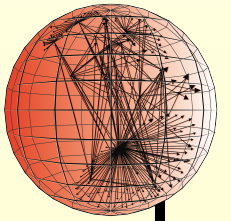
What is OKN-KISMET?

(first steps..)

Cutting edge Internet cartography measurement and analytic tools

Crucial operational network engineering expertise required for epistemologically sound interpretations of the measurements

Methodologies to combine different sources of data to reveal insights, and technology to responsibly manage data integrity, availability, and privacy.



Who is OKN-KISMET

primary investigators:

UC San Diego
JACOBS SCHOOL OF ENGINEERING
Computer Science and Engineering


PI: Kimberly Claffy


PI: David Clark


PI: Geoff Voelker

partners:

Massachusetts Institute of Technology

KO TE TANGATA

caida

NSRC
Network Startup Resource Center

NLNETLABS

University of Twente
The Netherlands

collaborators:

Penn Law
University of Pennsylvania Carey Law School

RIPE NCC
RIPE NETWORK COORDINATION CENTRE

Open Tech Strategies

GEORGETOWN LAW

ILLINOIS
UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN

DNS-OARC
Domain Name System Operations Analysis and Research Center

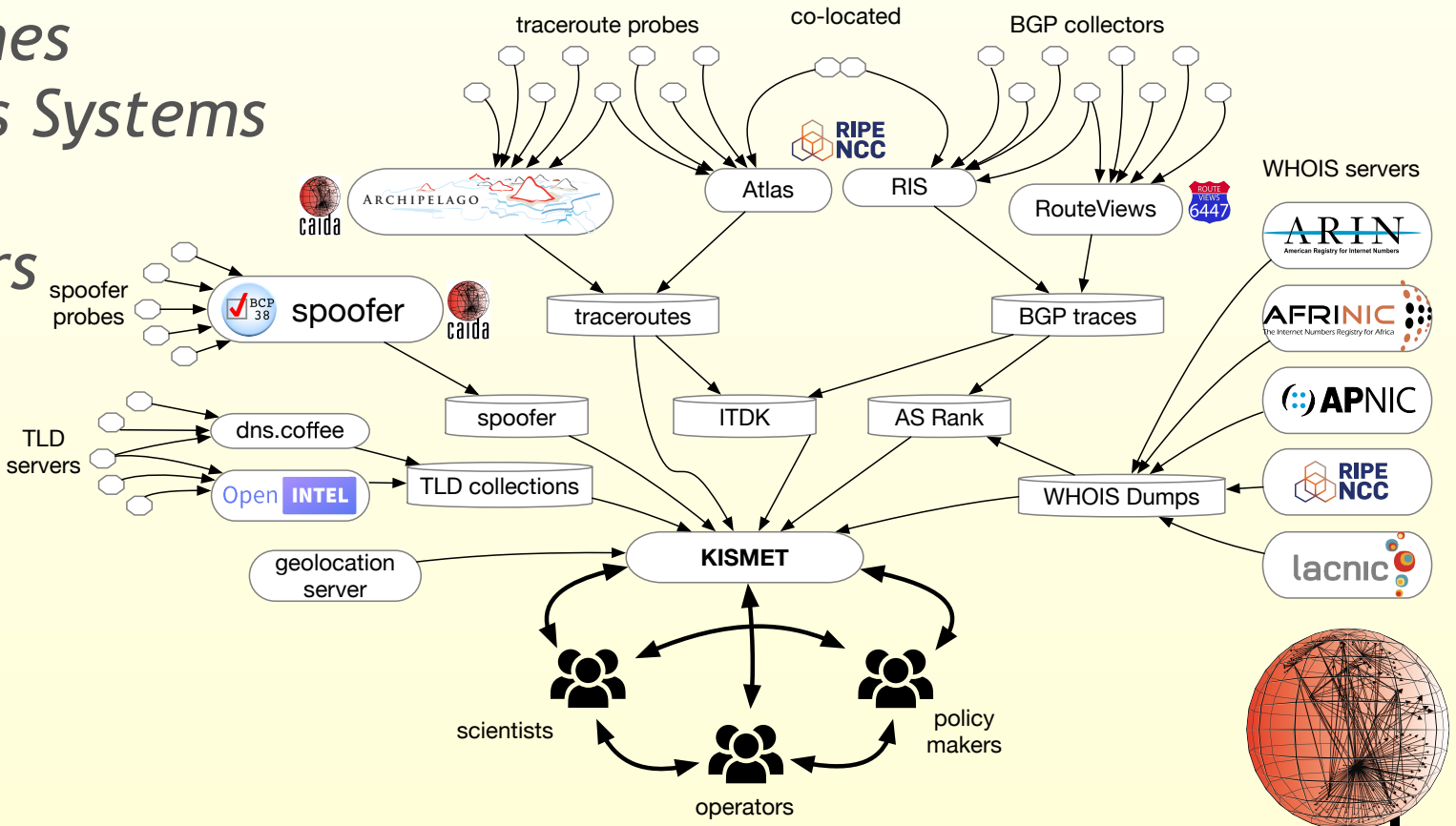
Tufts
UNIVERSITY

BRIGHAM YOUNG UNIVERSITY
FOUNDED BYU 1875
PROVO, UTAH

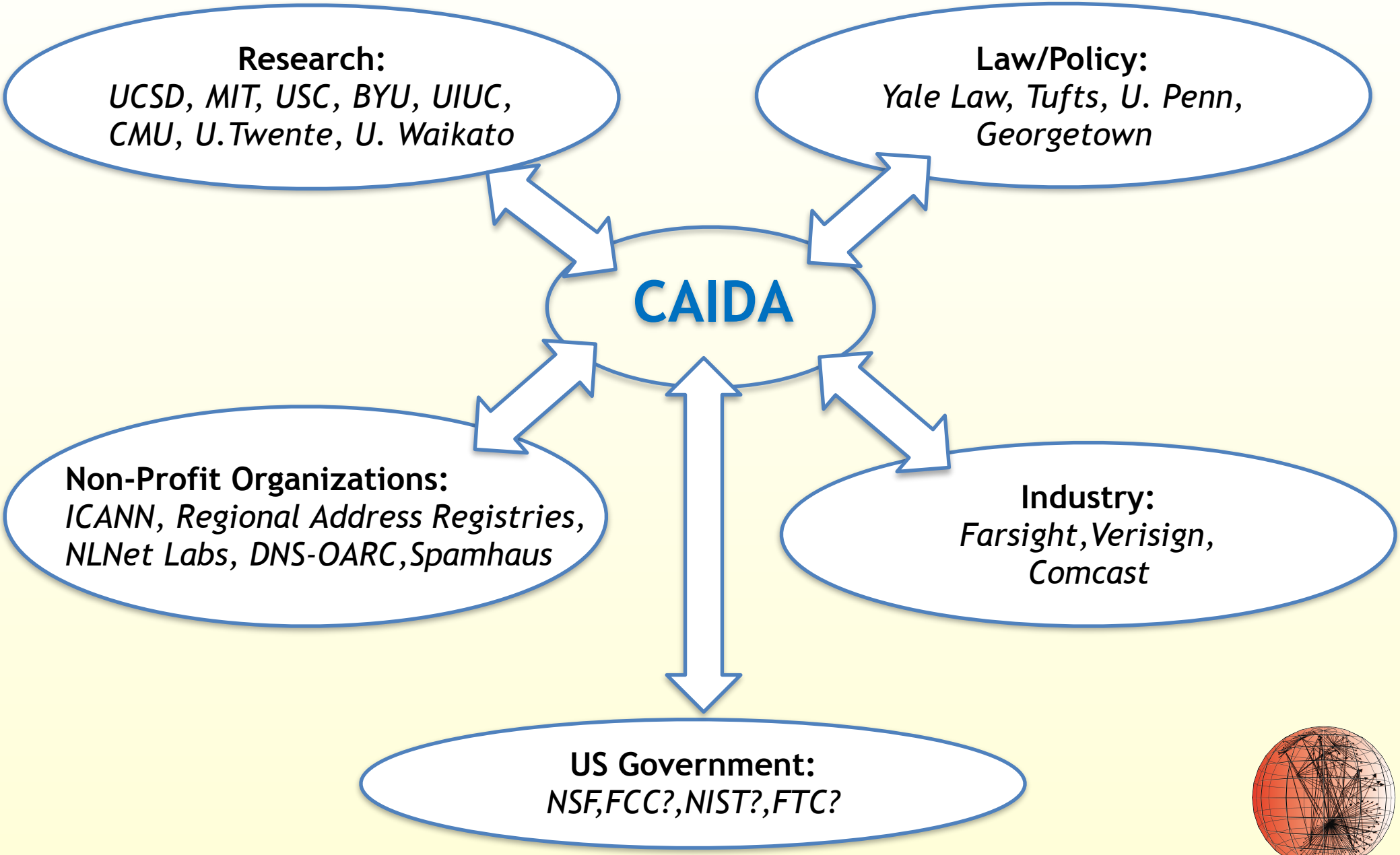
OKN-KISMET at a Glance

http://www.caida.org/funding/okn-kismet/

- **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*



KISMET Prospective Collaborators



caida

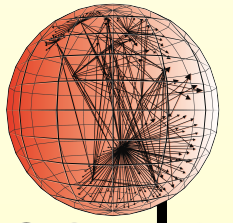
Enabling infrastructure research

Phase I-II+: Patterns, Protocols, Production

Study topics we want to enable in future (2+ years)

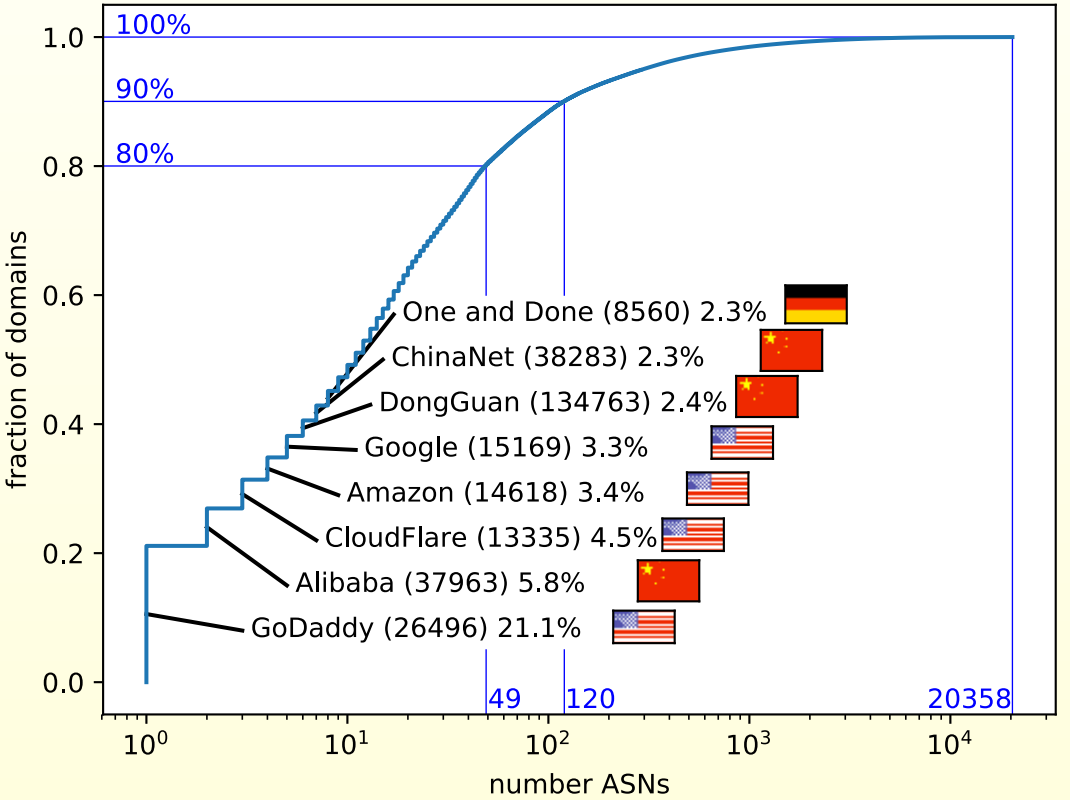
- Heavily annotated map of the namespace
- Knowledge graph to support pattern detection
- Combine DNS resolution and AS-level topology graphs. (who provides transit for various domains)
- Trends in concentration of domain hosting/name service and address space ownership/control
- Correlations of IP and DNS structure with blacklist and spoofing data

Support emerging (and struggling) movement for reproducibility



Representation Knowledge Graph (Phase II)

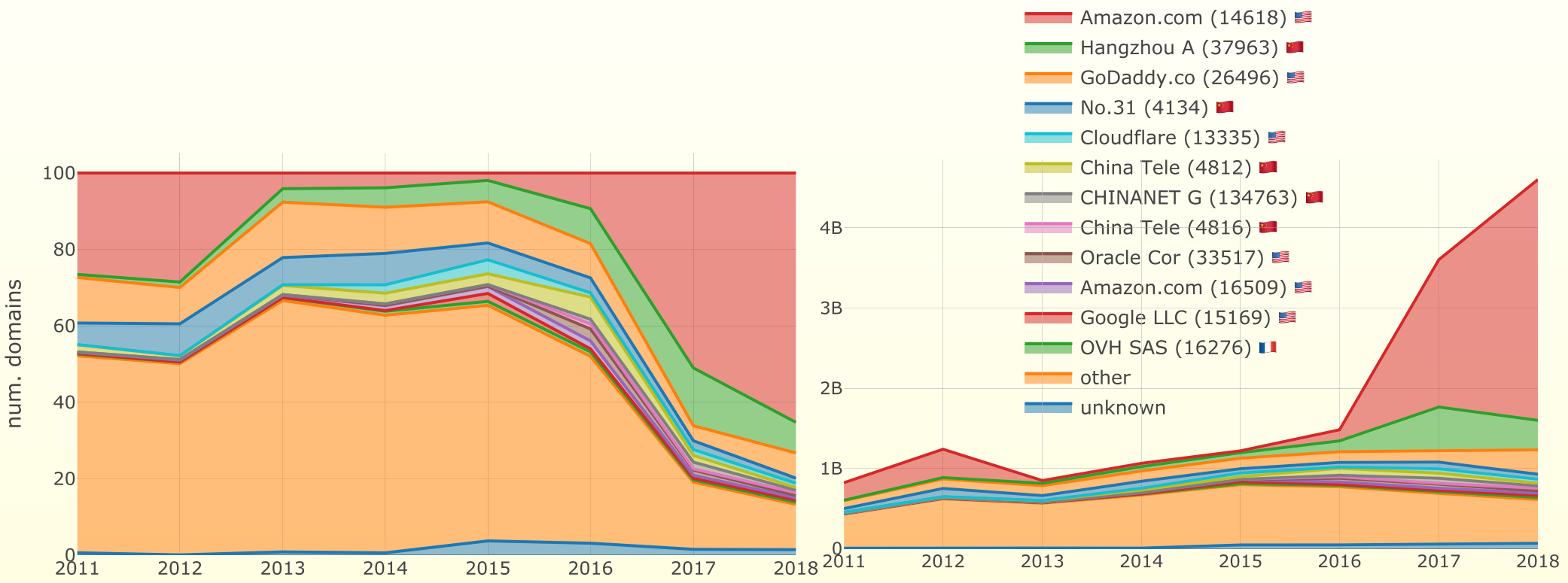
- Analysis and visualization modules -- to detect anomalous (suspicious) patterns
- Capture concentration of domains across registrars that are also autonomous systems



Concentration of second-level domains by registrar (1 May 2019 CZDS TLD zone files)

caida

Exploring DNS patterns

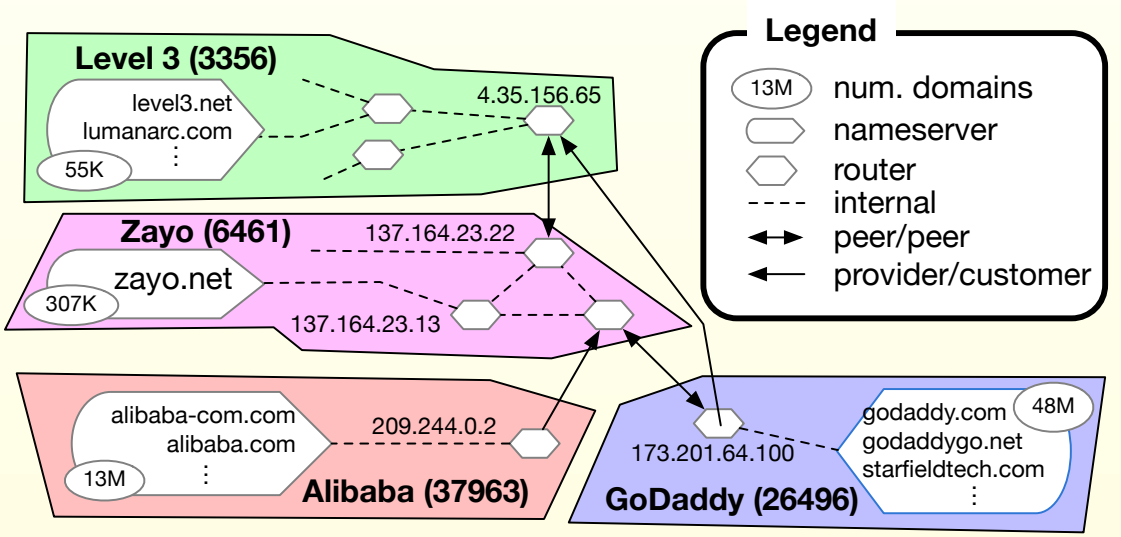


Starting in 2016, domain servers in Amazon's address space went from only 9.6% of to 65.2% of domains by adding 3 billion new domains. Is it Amazon or AWS customers?

caida

Superimposing the DNS Resolution and AS-level Topology Knowledge Graphs (Phase II if funded)

- Reveal ISPs providing hosting/transit for specific domains and TLDs
- Reveal physical co-location facilities
- Geolocation of IPs/ASNs



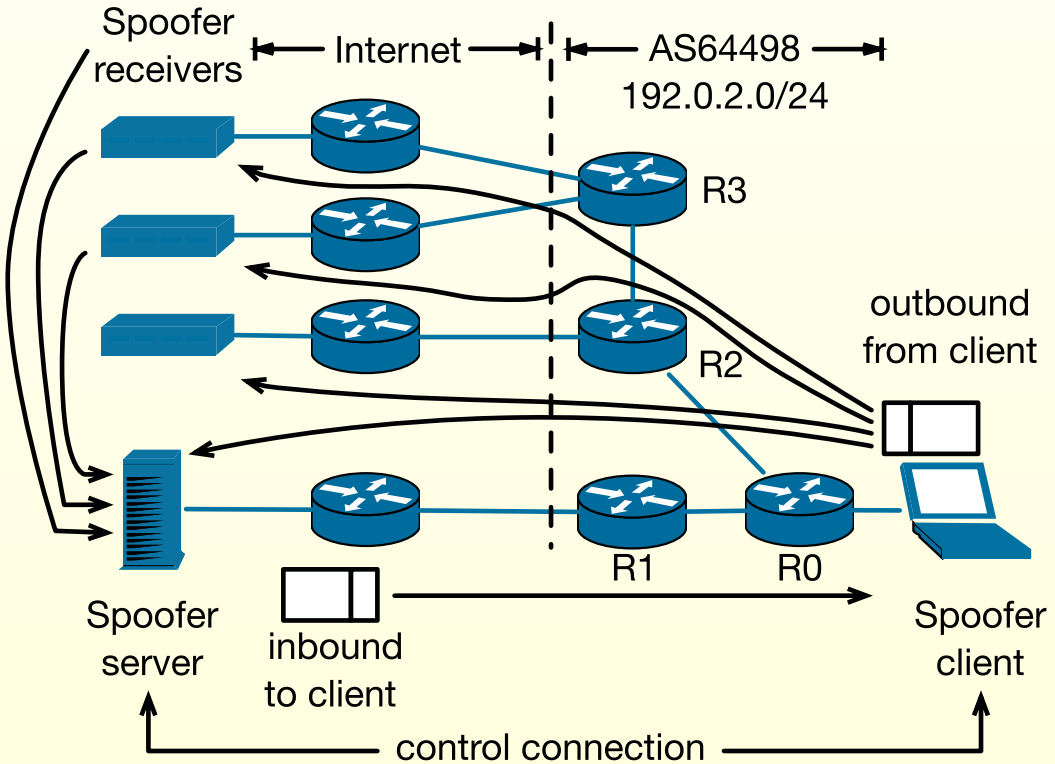
Capturing (simplified) DNS semantics an AS-level entity relationship network

caida

Possible Data Sources: Security Hygiene Data

http://www.caida.org/publications/papers/2019/network_hygiene_incentives_regulation/

CAIDA platform for studying deployment of Source Address Validation



Spoofer architecture:

- client software
- server coordinates measurements
- receivers collect client's packets

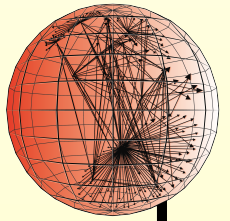
Cited in recent NIST recommendations on Securing Internet traffic.

caida

Potential KISMET-enabled R&D agenda for routing security

Can an OKN catalyze the scientific advancement of data processing techniques for hijacks detection and classification and protection?

- No shortage of data about the routing system
- Persistent knowledge gap
- No consensus on prevalence and impact of route hijacking attacks
- Need to derive knowledge from the measurements
- Need to use the knowledge strategically to improve security of routing system



Extending MANRS+ (BGP)

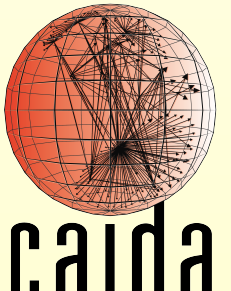
https://www.manrs.org/

MANRS project (current)

- **stated goal is to prevent route hijacking**
- **currently do not provide measurements or transparency**

MANRS+ (proposed)

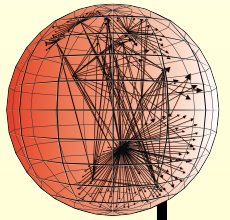
- **MANRS project should**
  - *Verify that a participating ISP meets all of its commitments*
- **ISPs**
  - *Flag dubious announcements by peers*
  - *Provide BGP monitor feeds to RouteViews or RIS*
  - *Validate customer and customer's announcement*



KISMET R&D agenda for DNS security

Need quantitative baseline description of many aspects of the DNS to show deterioration of the ecosystem.

- Knowledge graph of relationships among domains, registrars/registries, transit topology, geography
- Empirical understanding of utility of different blocking controls
- Map out control flows of DNS ecosystem (e.g., DOH)
- Map out money flows of DNS ecosystem
- Possibility (likelihood?) of avoiding DNS and use URL+IP
- Explore/model potential future scenarios



KISMET: Identified Needs

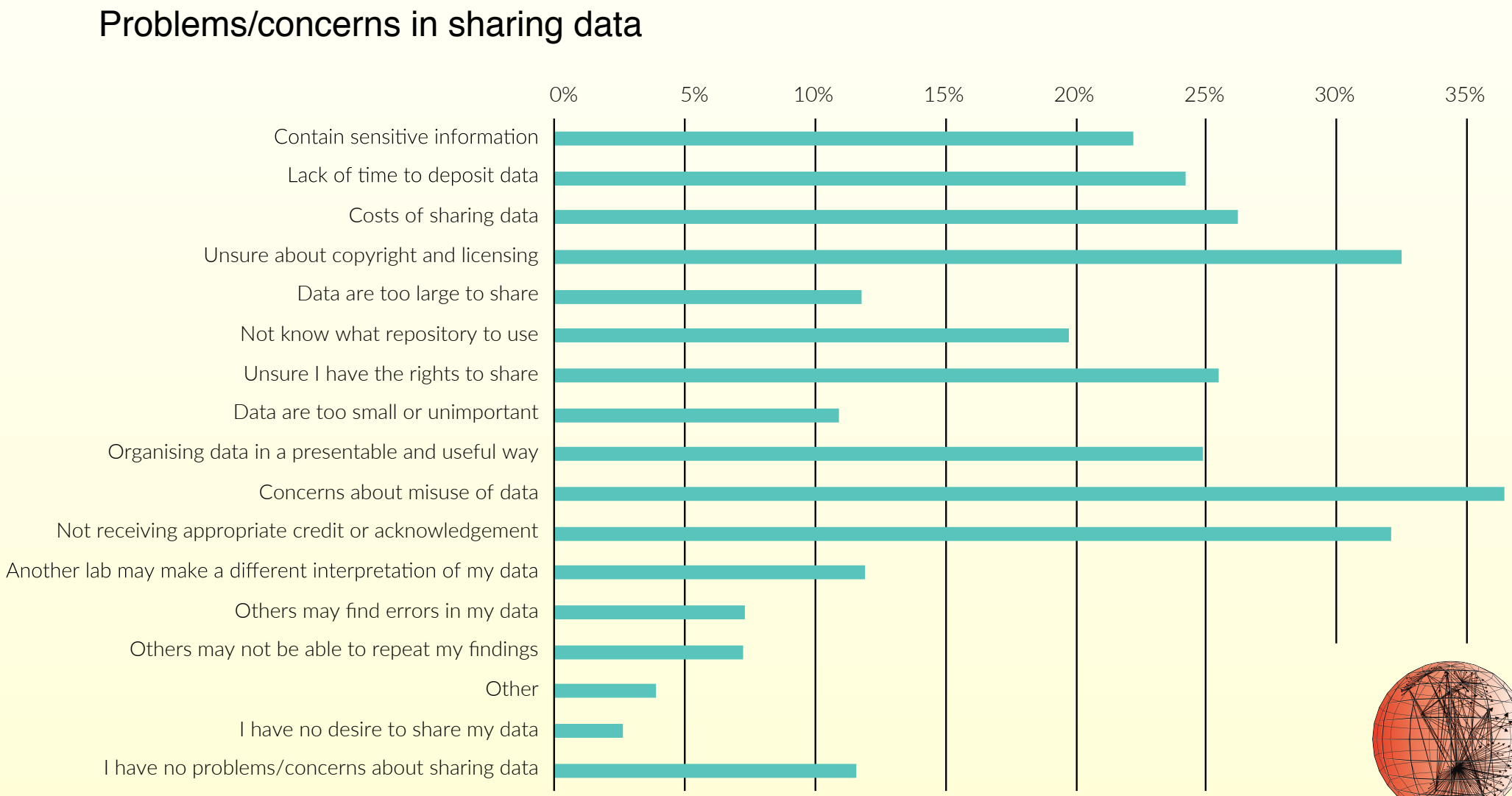
- Need to follow FAIR principles for Internet data management and stewardship
- Need for Standards
- Need for Meta-data that enables discoveries
- Need for Knowledge Graphs
- Need for Ontologies
- Need for Interoperability

Strategic role for NIST in these areas

caida

OKN Challenges

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

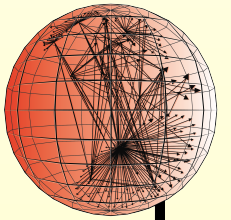


caida

Possible Discussion Areas

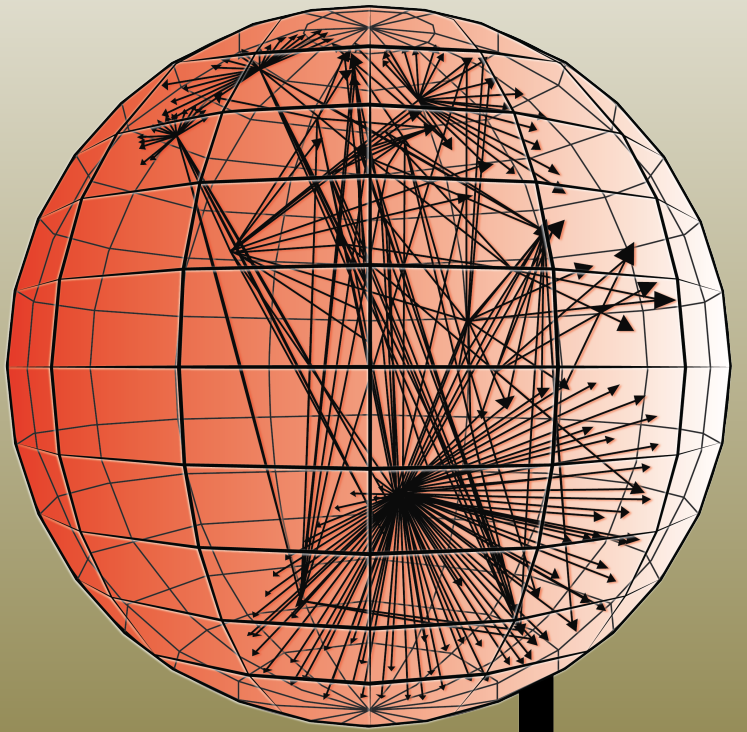
(NIST programs)

- Advanced DDoS Mitigation Techniques
- Comparison of Internet Congestion-Control Algorithms
- High Assurance Domains
- Internet Infrastructure Protection
- NIST Cybersecurity for IoT Program
- Robust Inter-Domain Routing
- Measurement Science for Complex Systems
- Security for Internet Systems
- Software Defined Virtual Networks
- Understanding Internet Performance from the User Perspective



Contact Information

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



caida