Homeland Security

Science and Technology

SECURING YOUR CYBER FUTURE

2016 | Cyber Security Division

R&D SHOWCASE AND TECHNICAL WORKSHOP

February 17–19, 2016

Washington, DC

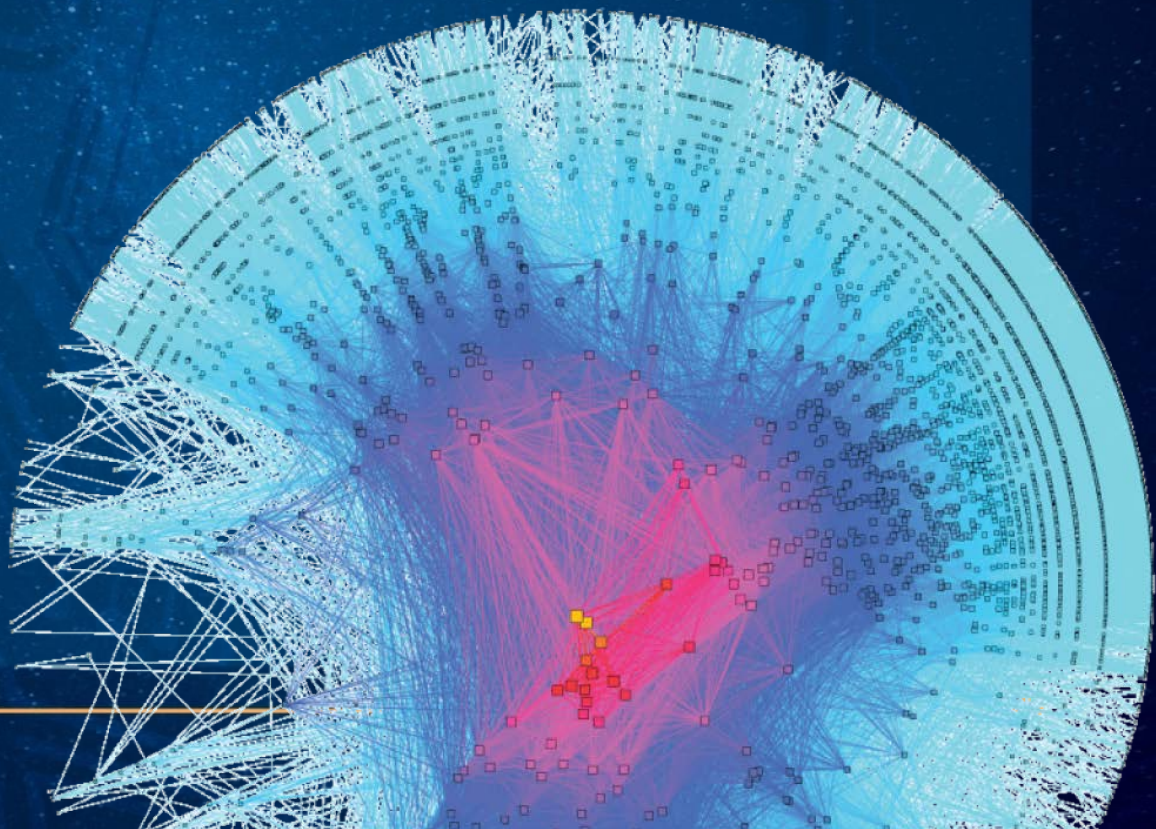
Homeland Security
Science and Technology

2016 | Cyber Security Division
R&D SHOWCASE AND TECHNICAL WORKSHOP

Cartographic Capabilities for Critical Cyberinfrastructure ("C4"): Internet topology and performance analytics for mapping critical network infrastructure

CAIDA/UCSD
PI k claffy

19 February 2016

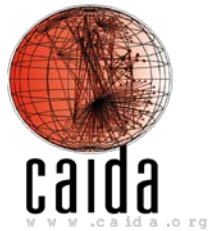


Team Profile

The Center for Applied Internet Data Analysis (CAIDA)

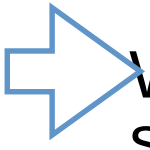
- Founded by PI and Director k claffy
- Independent analysis and research group
- 15+ years experience in data collection, curation, and research
- Renowned world-wide for data collection tools, analysis, and data sharing
- located at the University of California's San Diego Supercomputer Center

Key personnel: Bradley Huffaker, Young Hyun, Marina Fomenkov, Josh Polterock, Ken Keys, Matthew Luckie (now at Waikato), Amogh Dhamdhere, Vasilieos Giotsas



Needs: infrastructure protection, situational awareness

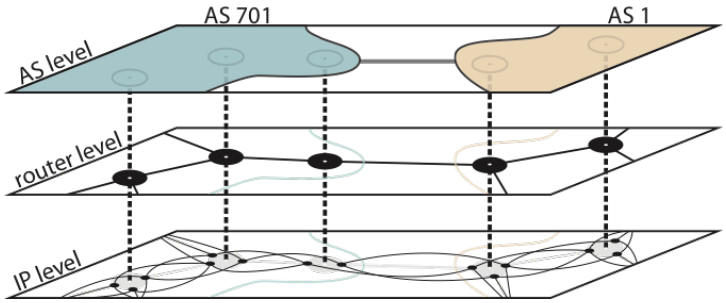
- No map of physical link locations, capacity, utilization, or interconnection arrangements.
- Best available data is incomplete and of unknown accuracy.
- Hinders efforts to: model network behavior and topology; design new protocols; assess security and stability properties
  - hygiene, robustness, resilience, and economic sustainability.



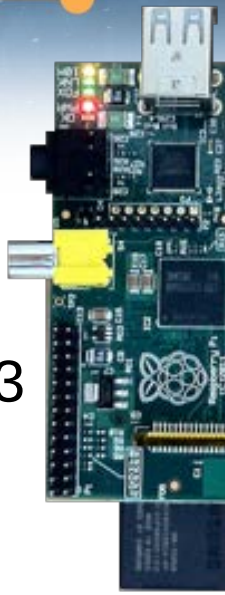
We designed, implemented, deployed, and operate a secure infrastructure that supports large-scale active measurement studies of the global Internet.

Motivation: (from DHS BAA)

- *"The protection of cyber infrastructure depends on the ability to identify critical Internet resources, incorporating an understanding of geographic and topological mapping of Internet hosts and routers. A better understanding of connectivity richness among ISPs will help to identify critical infrastructure. Associated data analysis will allow better understanding of peering relationships, and will help identify infrastructure components in greatest need of protection. Improved router level maps (both logical and physical) will enhance Internet monitoring and modeling capabilities to identify threats and predict the cascading impacts of various damage scenarios."*

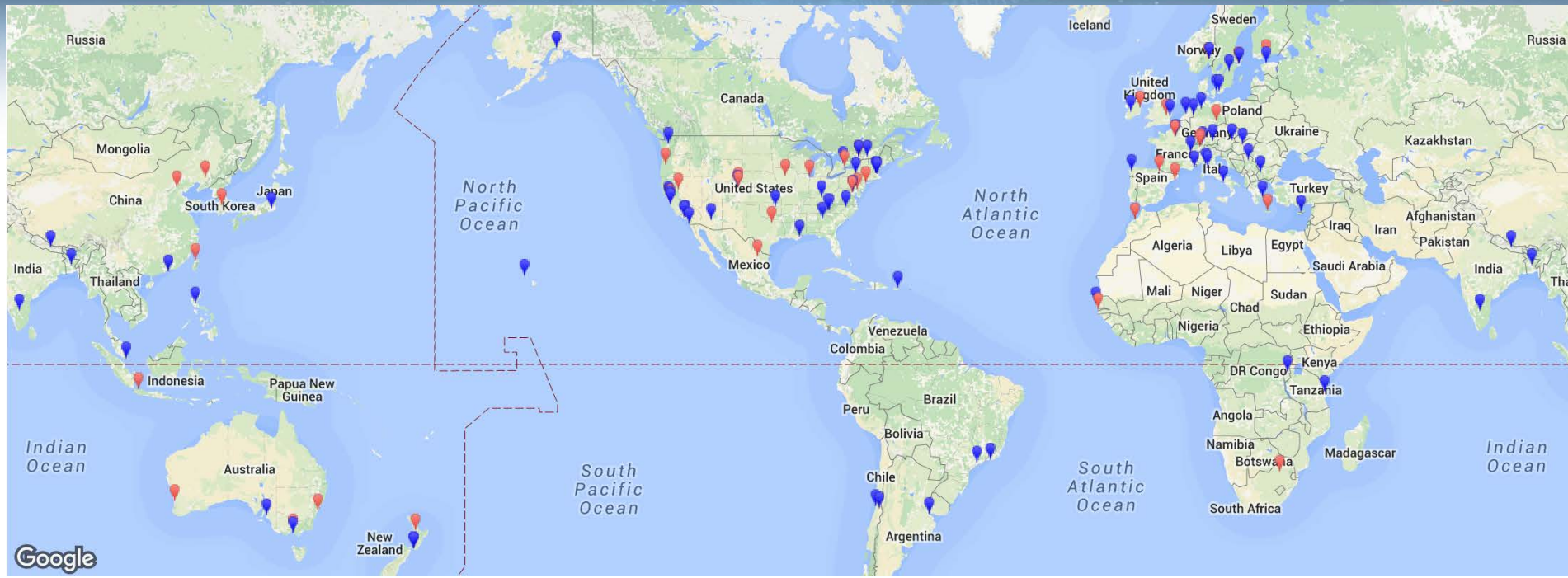


Approach: measurement science



- **Active measurement using Archipelago measurement infrastructure**
  - Continuous random probing of IPv4 address space
  - Tailored probes to elicit specific behavior, enable inferences
  - 137 monitors and growing (56 IPv6, 90 Raspberry PIs, 23 RadClock) — *ask us if you want one*
- **Improve completeness, accuracy, richness of map**
  - Incorporate other sources of data to expand coverage, visibility, and semantic labeling of map (BGP, traceroute, IXP, WHOIS, DNS, geolocation, traffic, economic)
  - Improved tools & methods for IPv4 & IPv6 alias resolution
  - unprecedented levels of validation
- **Support infrastructure hygiene assessments**
  - Vulnerabilities, routing stability, filtering hygiene, congestion
  - other experiments at http://www.caida.org/projects/ark/

Monitor Deployment



Legend: Raspberry Pi - FreeBSD

- 137 monitors in 44 countries
  - 90 Raspberry Pi's
  - 56 have IPv6
  - 23 have RADclock

Continent

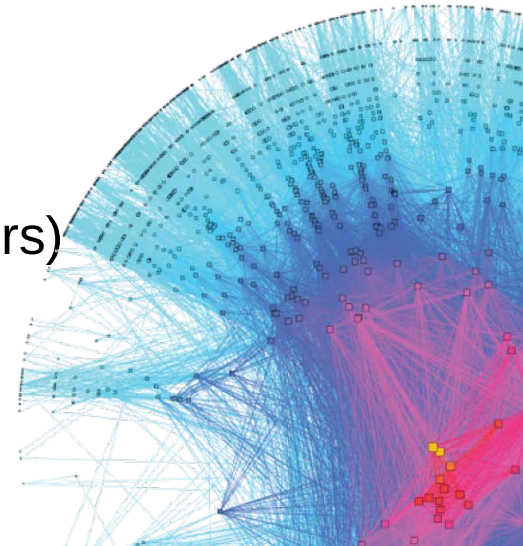
50	North America
7	South America
49	Europe
7	Africa
16	Asia
8	South Pacific

Organizations

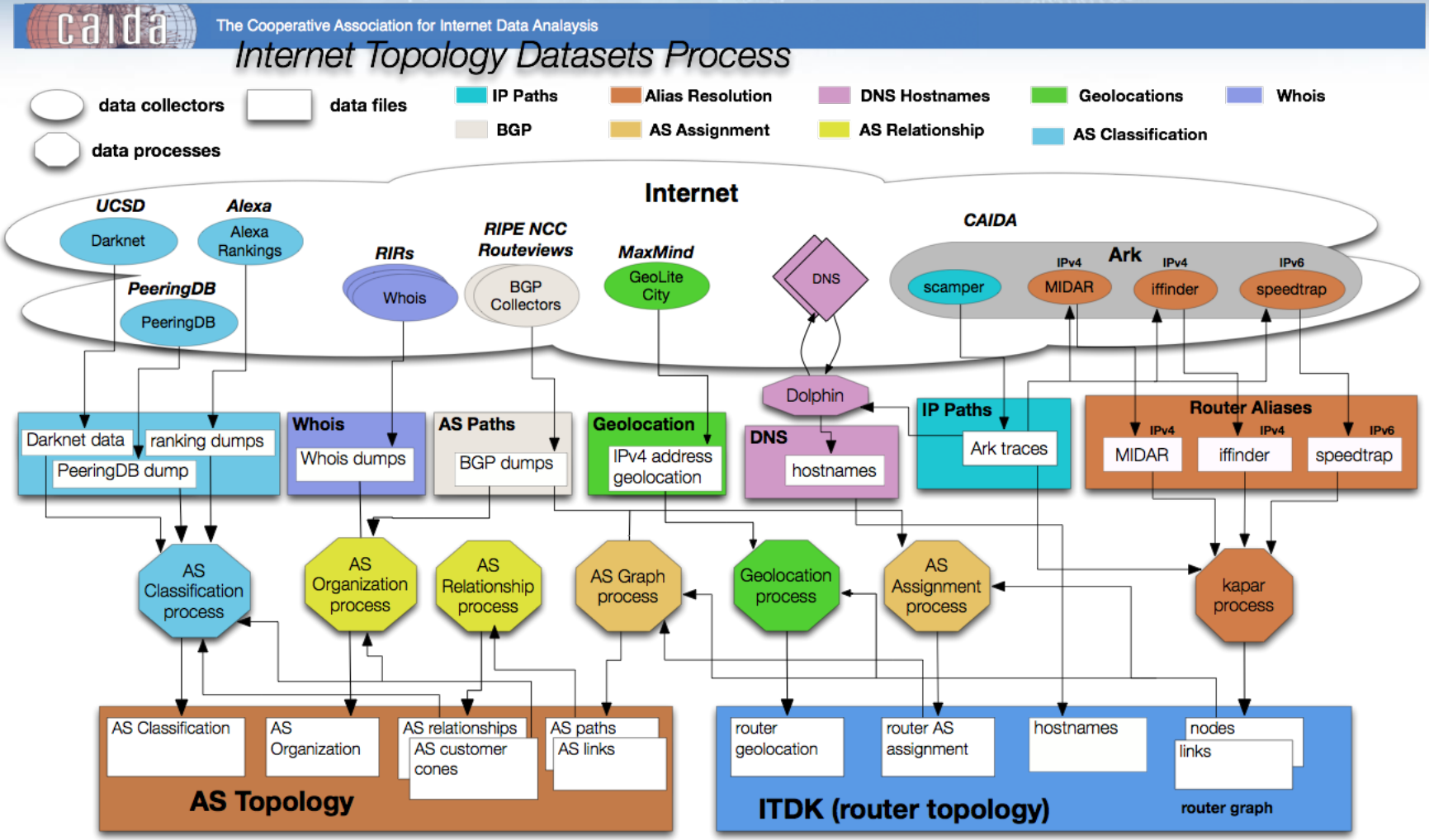
43	academic
59	residential
24	commercial/business
17	network infrastructure

Outcomes: Datasets

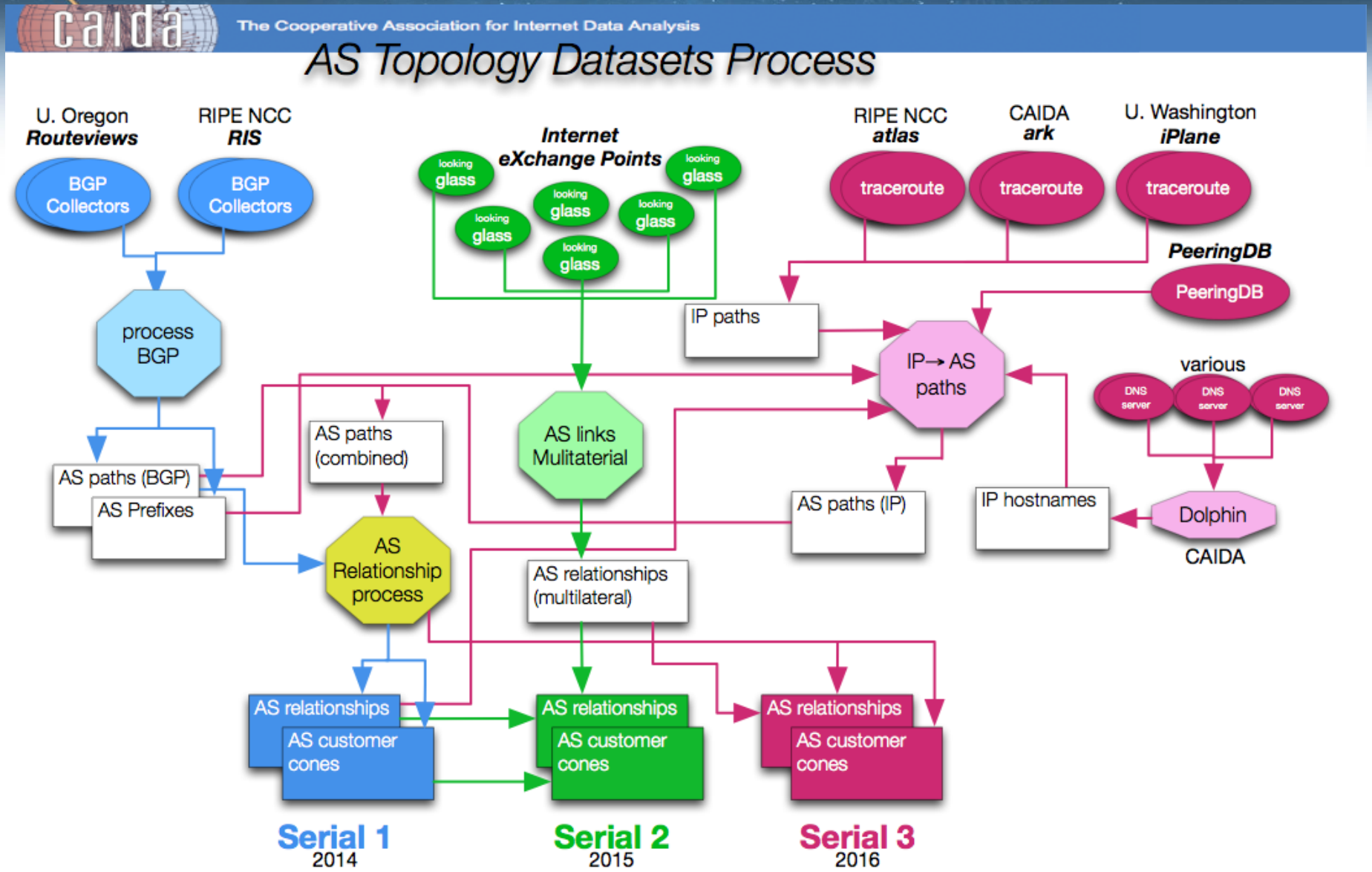
- **Community Source for IPv4/IPv6 Topology Datasets**
- **Web-based Datasets**
  - –Ark statistics http://www.caida.org/projects/ark/statistics/
  - –AS Rank http://as-rank.caida.org/ Interactive
- DNS Decoding Database (DDec) http://ddec.caida.org/ Interactive
- **On-demand Measurements**
  - tod-client (topology on-demand)
  - Vela.caida.org: Web Interface
- **Developing Experimental Capabilities**
  - Outage detection (e.g. cable cuts, natural disasters)
  - BGP hijacks
  - Interconnection Congestion
  - *<your experiment here>*



Internet Topology Datasets Process



AS Relationships: Process



Current Status: Resulting Science

- **Papers**
  - "Internet-Scale IPv4 Alias Resolution with MIDAR" (ToN13)
  - "AS Relationships, Customer Cones, and Validation" (IMC13)
  - "Inferring Multilateral Peering" (CoNEXT13)
  - "A Second Look at Detecting Third-Party Addresses in Traceroute Traces with the IP Timestamp Option" (PAM14)
  - "DRoP: DNS-based Router Positioning" (CCR14)
  - "Spurious routes in Public BGP Data" (CCR14)
  - "Challenges in Inferring Internet Interdomain Congestion" (IMC14)
  - "Inferring Complex AS Relationships" (IMC14)
  - "Measuring and Characterizing IPv6 Router Availability" (PAM15)
  - "IPv6 AS Relationships, Clique, and Congruence" (PAM15)
  - "Resilience of Deployed TCP to Blind Attacks" (IMC15)
  - "Mapping Peering Interconnections at Facility Level" (CoNEXT 15)
- **Community** support**:** hosted **AIMS 2013, 2014, 2015, 2016 (Feb)** (http://www.caida.org/workshops/aims) all reports published in CCR

Current Status: Resulting Science

"Mapping Peering Interconnections at Facility Level" (CoNEXT15 Best Paper)

- **constrained facility search** to infer physical interconnection
- relies on published data about networks at facilities
- traceroutes from > 8500 servers around the world to identify interconnection engineering strategy
- many routers implement private & public peerings, via multiple IXPs
- engineering strategy inference constrains set of facilities such that one can often identify specific facility where a given interconnection occurs

Current Status: Resulting Science

"Resilience of Deployed TCP to Blind Attacks" (IMC15 Best Paper)

- **infrastructure hygiene assessment:** how many deployed TCP stacks are vulnerable to blind in-window attacks? (experiment ran Sept 2015)
  - off-path adversary disrupts a connection by sending a packet that the victim believes came from its peer, causing data corruption or connection reset
  - finding: *38.4% vulnerable to at least one of three in-window attacks we tested*. router vulnerabilities worse
- **supports case for systematic, scientific, longitudinal empirical analysis of critical infrastructure**
  - and for better mechanisms to incent security hygiene

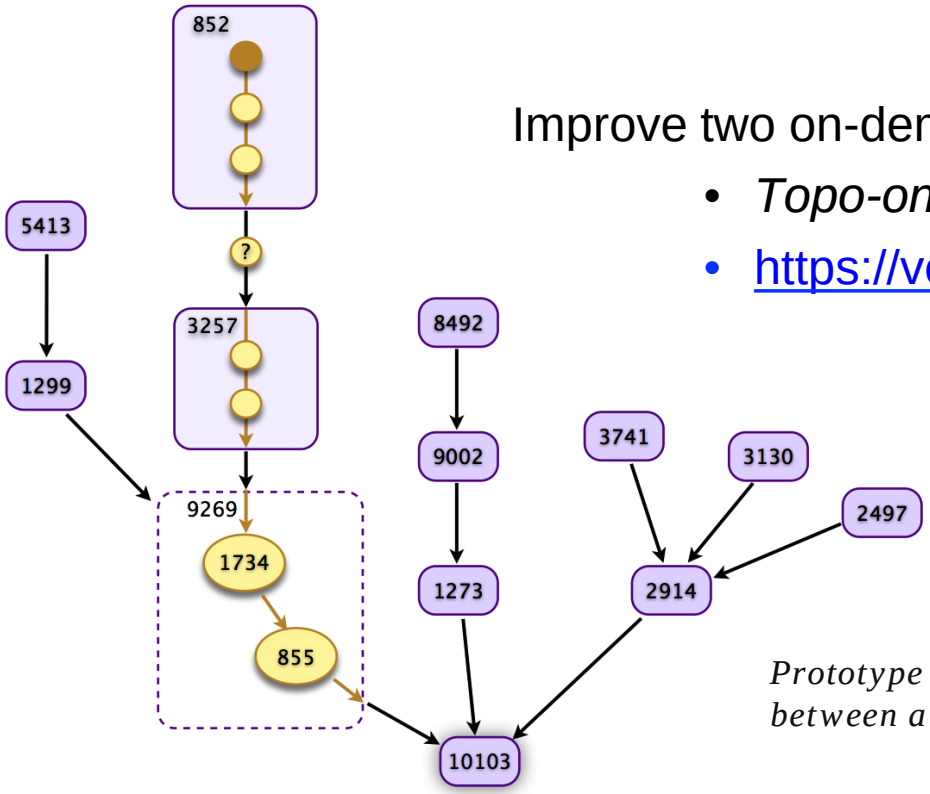
Benefits

Improved situational awareness of the Internet through:

- **Increased completeness**
  - Increased measurement infrastructure
  - Expanded and more efficient probing
  - New methods to synthesize disparate Internet topology data
- **Increased accuracy**
  - Filter out (some) false link inferences, assess impact
  - Improve AS business relationship inference
- **Improved richness of topology maps**
  - Better geolocation accuracy
  - Router level: aliases resolved w/2 methods (min FP or max coverage)
  - Increased connectivity at router-level
  - Physical facility awareness
  - IP, router, PoP, and AS-level
  - AS-level annotations: org, type, relationship, performance

Next Steps (Data Accessibility)

Create an interface for **browsing**, **querying**, and **visualizing** the data gathered by the infrastructure.



Improve two on-demand topology measurement tools

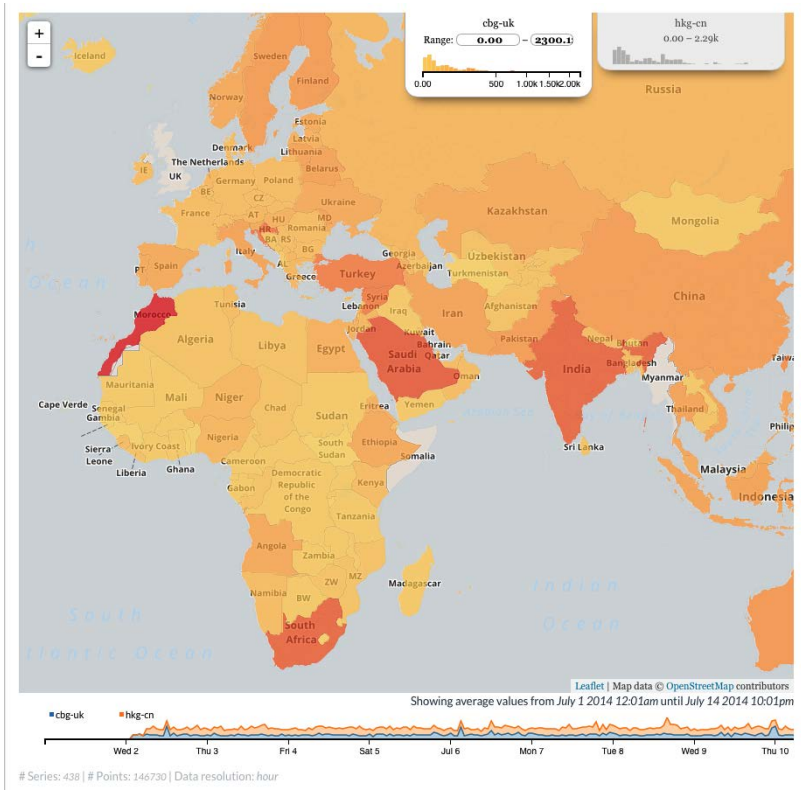
- *Topo-on-demand* – CLI to Ark platform
- https://vela.caida.org/ web GUI to Ark platform

Prototype visualization showing differences between a traceroute path and BGP AS paths

Next Steps (Data Accessibility)

- *browsing* interface
  - view broad properties and summary statistics over multiple time scales and aggregation levels
    - example: trace counts and response rates; path-length and RTT distributions; inferred AS links

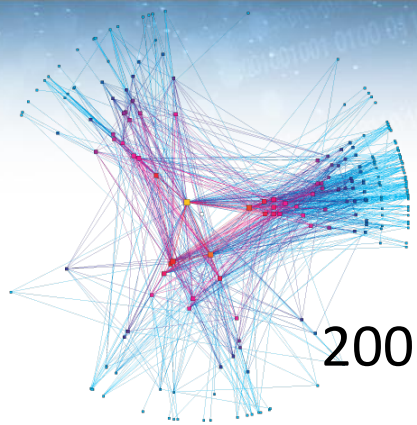
Prototype view of traceroute RTTs implemented with CAIDA's Charthouse



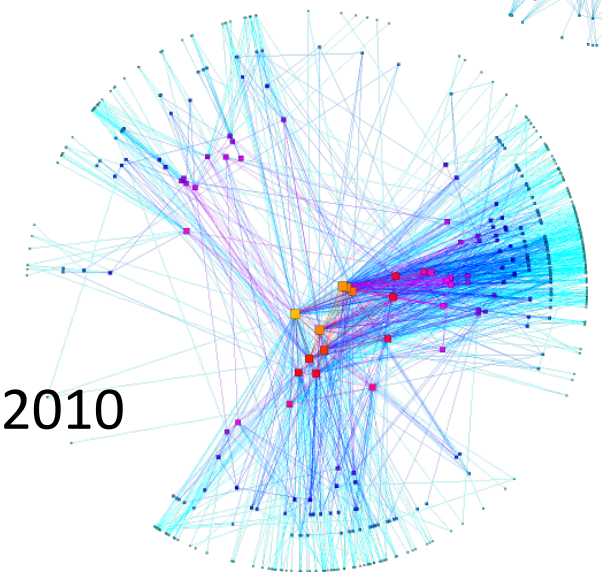
Next Steps (Data Accessibility)

IPv6 appears to finally being on it's way to a real network!

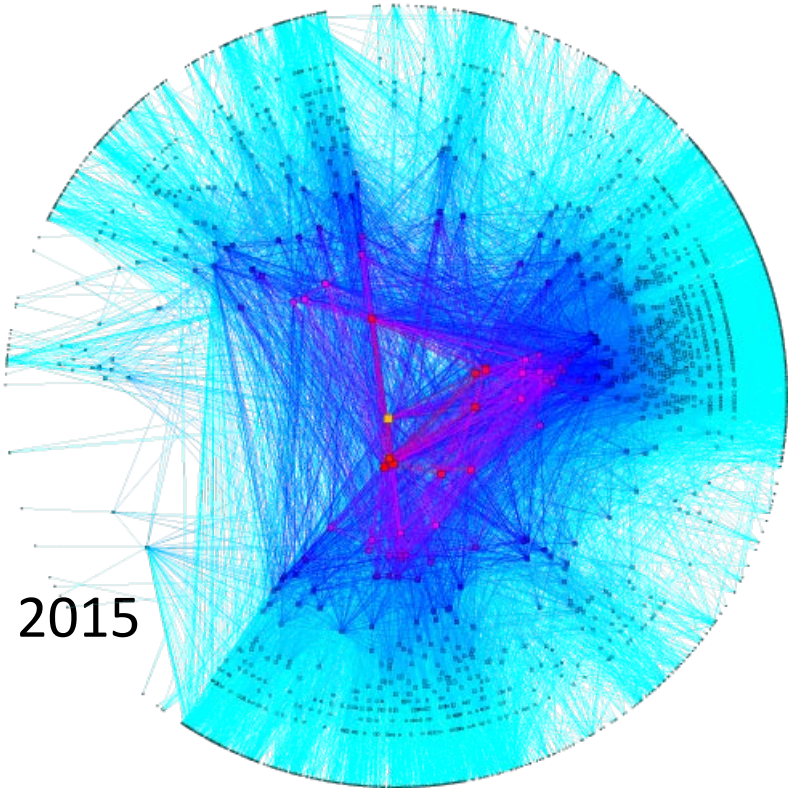
IPv6 AS Core



2005



2010



2015

Next Steps (Data Accessibility)

- *query* interface
  - find the most relevant historical data for one's research
    - either directly answers a question, or identifies data to download for further study

  examples:
    - all traceroutes through a given region and time period toward/across a particular prefix/AS/[*country?]
    - router address aliases for a given IP address
    - all inferred links to a router identified by a given IP address
    - all routers in a given city*

*[*blocked on improved geolocation of routers]*

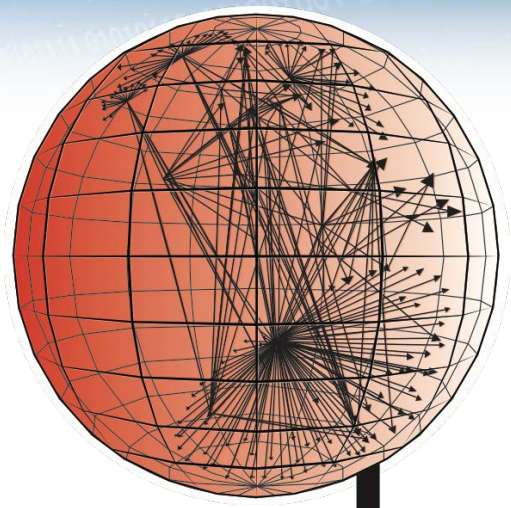
~~Competition~~ — Related Work

- RIPE Atlas (http://atlas.ripe.net/)
- Internet Atlas (http://internetatlas.org/)
- iPlane datasets (http://iplane.cs.washington.edu/data/data.html)
- zMap (https://zmap.io/), with results (https://censys.io)
- ISI Census (http://isi.edu/ant/address)

- Renesys (http://www.renesys.com/) recently acquired by Dyn

Contact Information

k claffy
kc@caida.org
CAIDA/UCSD
858-534-8333



caida

SDSC
SAN DIEGO SUPERCOMPUTER CENTER

UC San Diego

Homeland Security

Science and Technology

2016 | Cyber Security Division

R&D SHOWCASE AND TECHNICAL WORKSHOP