2017 Cyber Security R&D Showcase and Technical Workshop

July 11 - 13, 2017 | Washington, D.C.

Homeland Security

Science and Technology

Mapping our Way to a More Secure Internet

k claffy / CAIDA/UCSD

11 july 2017

Homeland Security

Science and Technology

Team Profile

The Center for Applied Internet Data Analysis (CAIDA)

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

Principal Investigators: kc claffy, Alberto Dainotti, Amogh Dhamdhere

Key Personnel: Bradley Huffaker, Young Hyun, Marina Fomenkov, Josh Polterock, Ken Keys, Matthew Luckie (now @Waikato), Vasilieos Giotsas, Paul Hick

Need

Today the "cyber threat" is one of our most serious economic and national security challenges.

But we lack understanding of the structure, dynamics, and vulnerabilities of the global Internet.

Measurement infrastructures, reliable, representative, Internet data sets, and advanced analysis tools are rarely available to researchers and developers.

Security First?

No, actually. Measurement first.

We cannot secure what we cannot measure.

Internet measurement & mapping is essential to cybersecurity

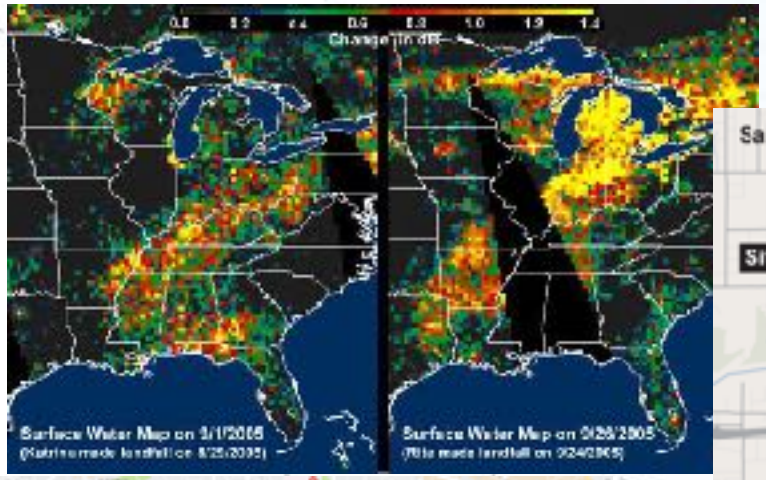
Outline of Talk

- What kind of Internet **maps** do we need? What **problems** do they solve?
- What **capabilities** are required to construct which maps?
- What are **challenges** to these capabilities? How many ways can we get the required data?
- Examples of **applied mapping** R&D
- How you can help!

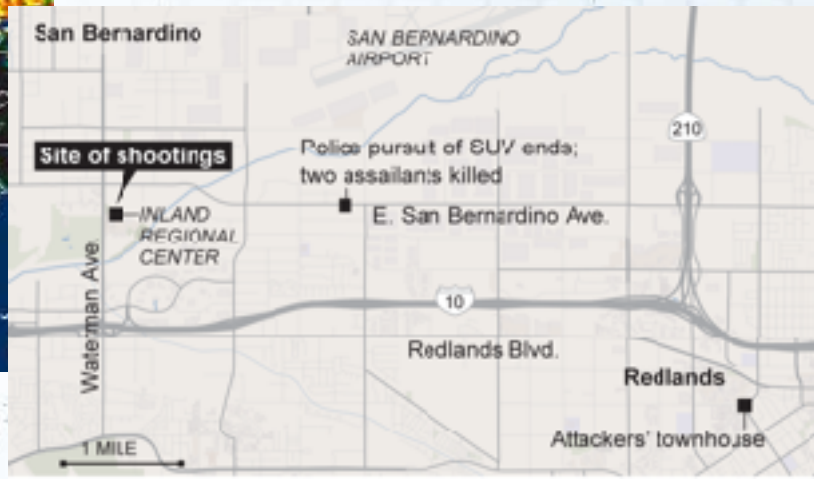
Real world crisis!

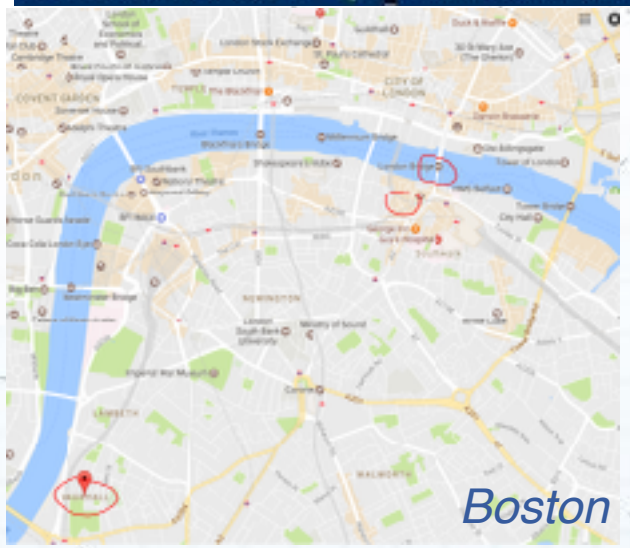
- **What do first responders do *first*?**
- **Find the crisis on a map!**
  - Where is it?
  - What is its scope?
  - What is its extent?
  - How fast is it moving?
  - How Do I Get There?

Katrina



San Bernardino, CA





Boston

Cybersituational awareness

How can maps inform understanding and mitigation of risk?

- ***physical*** links, interconnection facilities, routers
- structure, dynamics, ***outages***
- ***resiliency*** to upstream disruption
- ***attacks*** in progress, e.g., DDoS, hijacks, censorship
- address space utilization, reachability, ownership
- security ***vulnerabilities***

But map construction requires data,
…. map utility requires info visualization,
and map validation requires social interaction

Mapping is interdisciplinary

Science, Technology, Art, and Politics

- In the case of entirely new and dynamic territory, new methods and techniques for cartography
- Infrastructure to operationalize it
- Platforms to share, improve, apply its utility
- Data privacy policies to protect sensitive data
- Funding models to sustain it (and leverage capabilities more broadly)

Macroscopic assessments

Guiding principle: ***We cannot secure what we cannot measure.***

1. **Baseline path measurements** to support anomaly detection
2. **Router-level topology** (ultimately on-demand)
3. **Facility-level topology** (physical interconnection)
4. **Performance** (which may reflect security incidents)
5. **Security hygiene best practices**, e.g., spoofer

Many others: security protocol compliance, TCP vulnerabilities, address utilization census, grey-market address transfers, <your experiment here>

Internet cartography primer

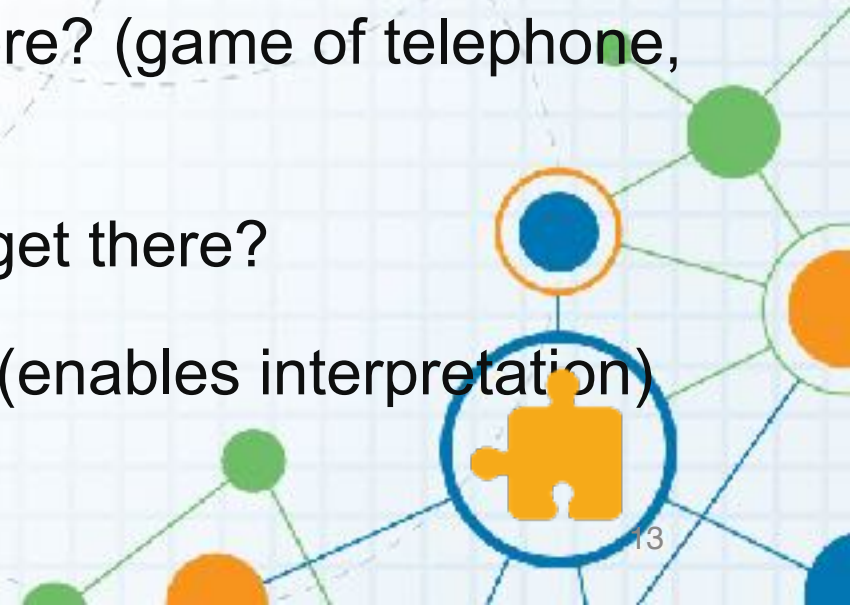
Guiding principle: ***The Internet was not designed to be measured.***

1. Internet changing every day
2. IP architecture does not respect (or acknowledge) network boundaries.
3. Basic topology measurement machinery is a 30-year old clever hack, with myriad misleading artifacts in output.
4. Network address assignment strategies and router implementation variance limit the accuracy of any single method
5. Many operators are not incentive-aligned to facilitate mapping research

Internet cartography primer

Baseline Internet measurement terms.

1. **Ping**: is this device (interface) responsive?
2. **Traceroute**: How does my packet get across the Internet? What is the IP-level path from here to there? (reported by intermediate IP hops, sometimes incorrectly)
3. **BGP**: what is the network-level path from here to there? (game of telephone, sometimes confusing)
4. **Round-trip time** (latency): how long does it take to get there?
5. **Metadata** galore: WHOIS, IANA, DNS, geolocation (enables interpretation)

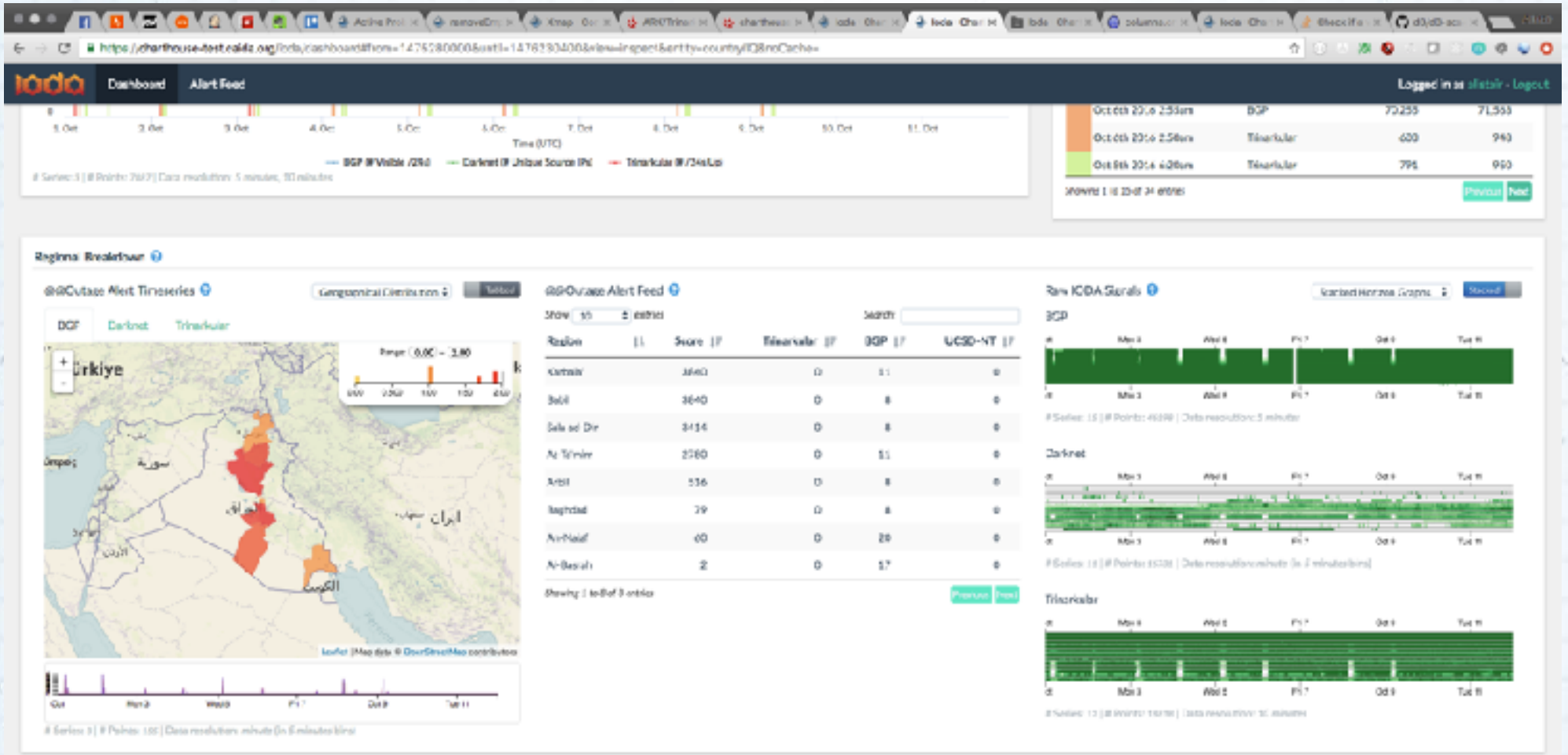


CyberCartography

- CAIDA has been developing Internet cartography techniques, and operating infrastructure to support these techniques, for over a decade
- Archipelago platform now 180 globally distributed nodes collecting data and performing experiments for researchers
- Multiple modes of access, from web interface to API to ssh access to nodes
- Now using platform to involve broader research community
- Linking to other platforms to amplify utility

[please go see demo at showcase!]

Internet Outage Detection & Analysis



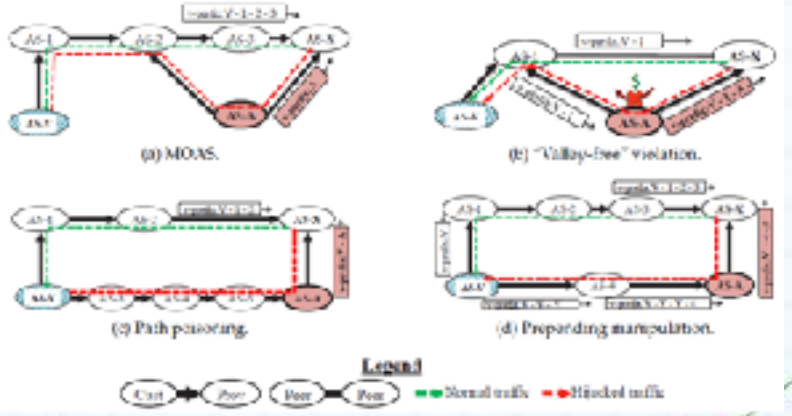
Screenshot of the IODA dashboard highlighting outages in Turkey

NSF CNS-1228994, UC San Diego
PI: Alberto Dainotti, CoPI: KC Claffy

To measure outages, we need a baseline

1. **Data acquisition and processing pipeline**
   a. sanitize BGP data from sliding 1-week of RV/RIS data (reconstruct routing tables of hundreds of operational routers worldwide with 5-min granularity)
   b. dynamically identify which IP addresses to probe (spatially and temporally fine-grained monitoring)
   c. Result: (nearly) live mapping of control and data planes
   d. Sharing: curated daily files of AS paths for post-event analysis
   e. Support: outage detection & analysis (IODA) system
   f. Support: interactive monitoring to detect route hijacks

Detecting traffic interception using real-time BGP anomaly detection paired with active probing triggered upon event detection

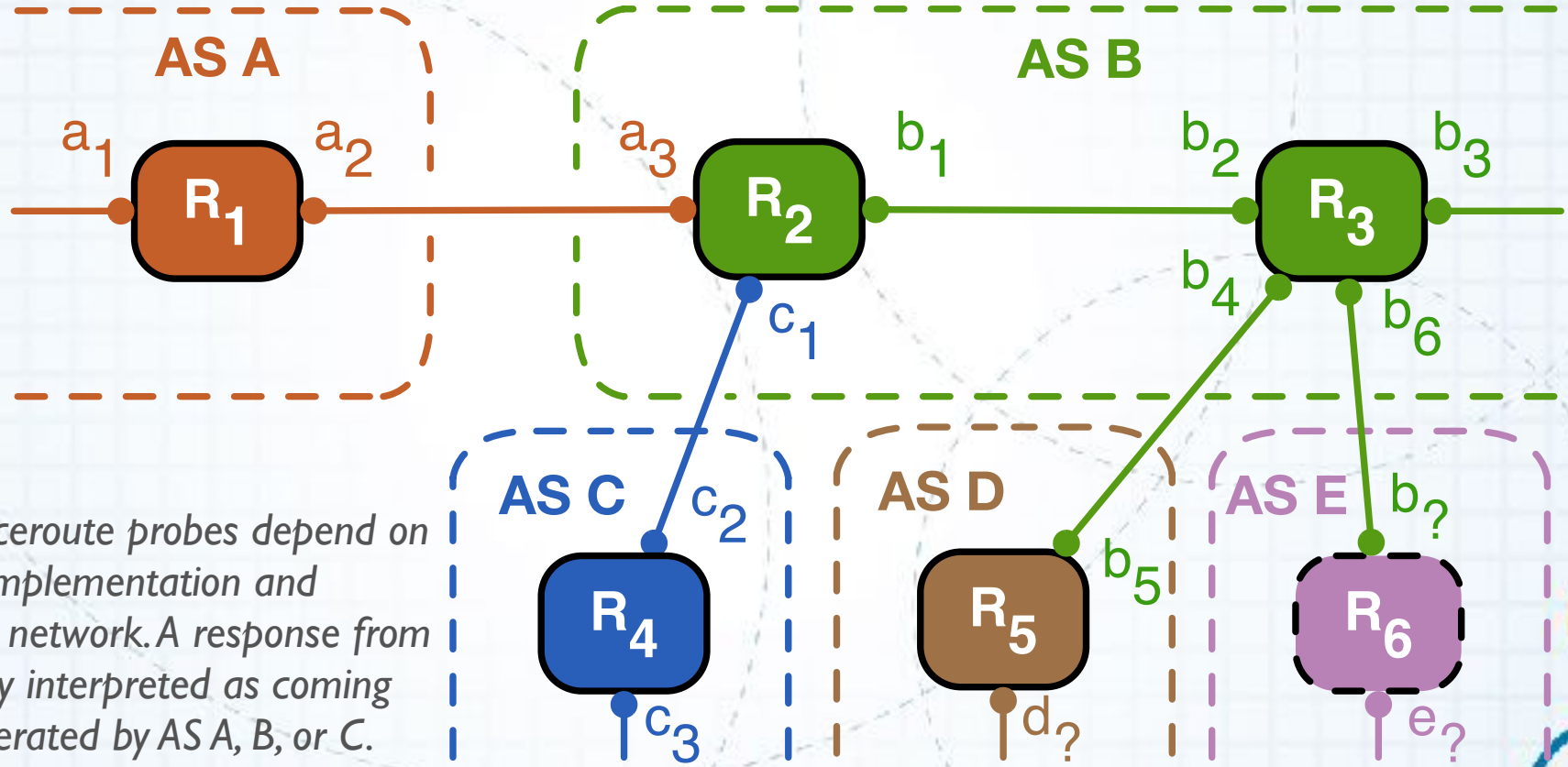


Mapping Interdomain Congestion

- goal: system to monitor interdomain links and their congestion state (can indicate DOS attack)
- near real-time “congestion heat map” of the Internet
- increase transparency, empirical grounding of policy debates

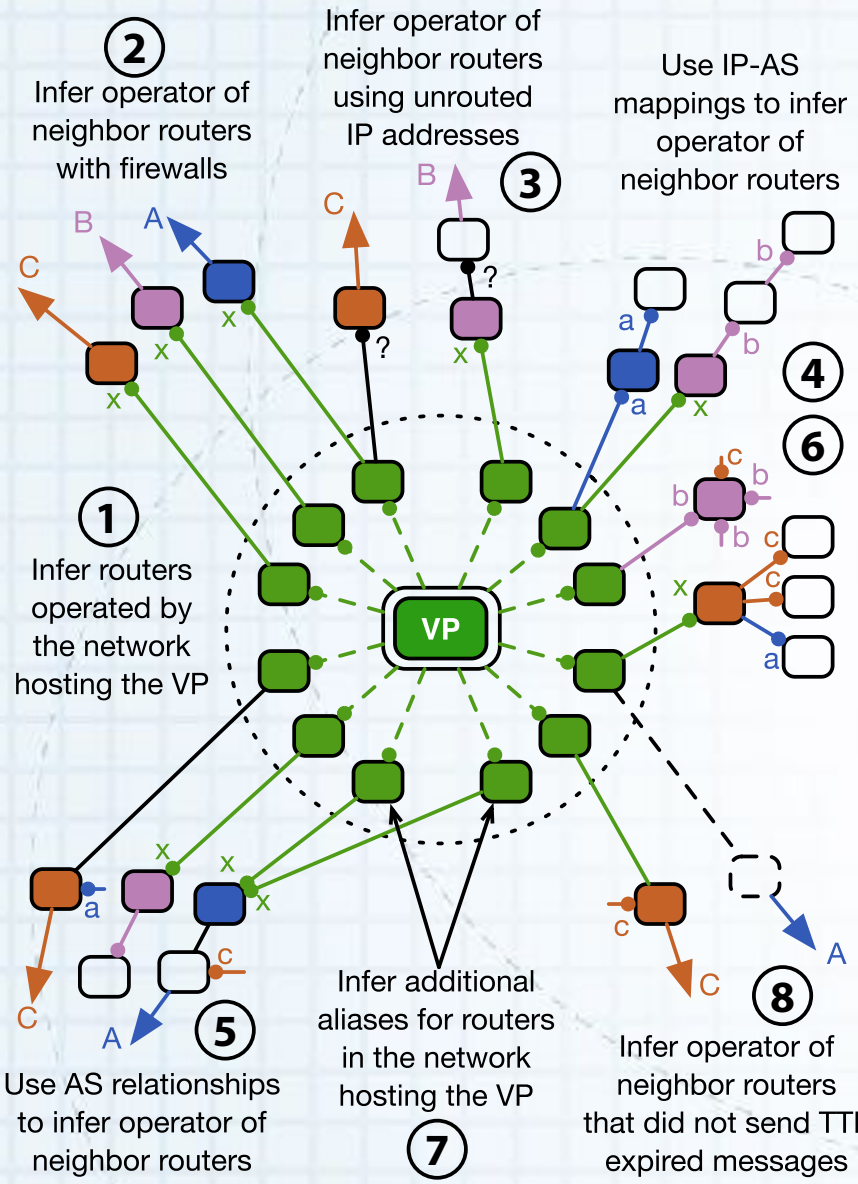
Mapping Router Interconnections

- Collaborating PI: Matthew Luckie (U. Waikato)



Responses to traceroute probes depend on router software implementation and placement in the network. A response from R_2 may be naively interpreted as coming from a router operated by AS A, B, or C.

Mapping interconnections

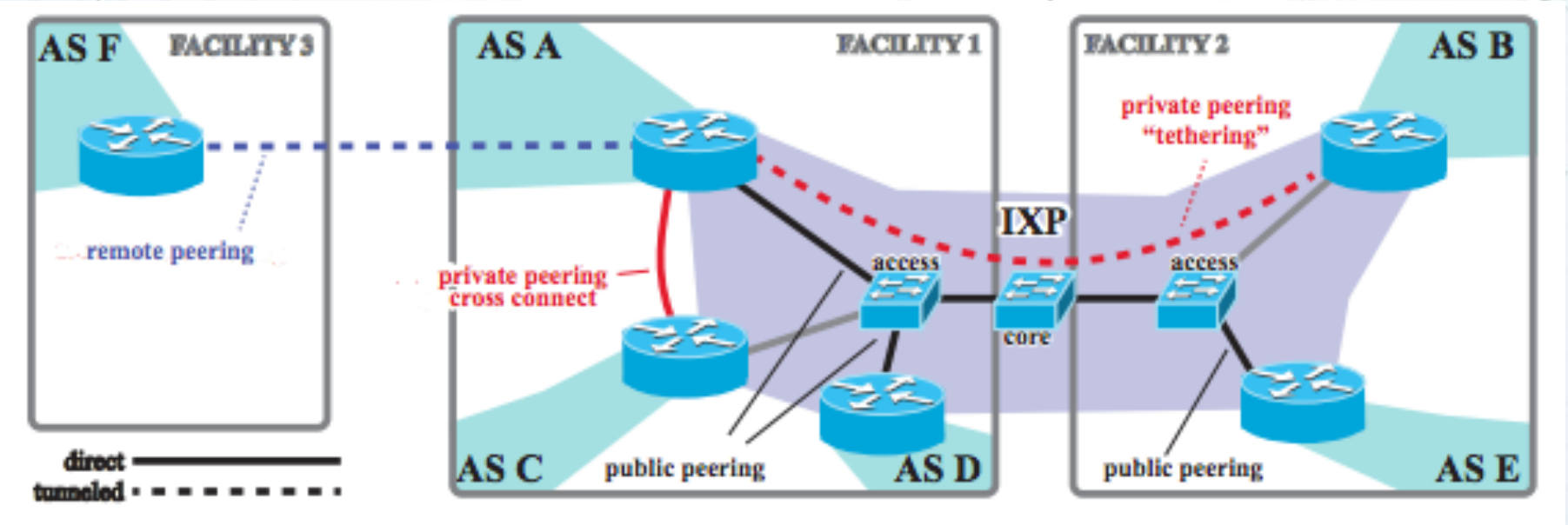


Conceptual mapping of heuristics to infer border routers (Diagram taken from Luckie, et al. IMC2016 paper.)

1. Developed and deployed heuristic algorithms to accurately infer router ownership & interconnections from traceroute data
2. Apply heuristics to annotate maps
3. Validation with ISPs
4. AS border mapping software runs continuously on Ark
5. Supports CAIDA's MANIC platform (Measurement and Analysis of Interdomain Congestion) [see demo]

Mapping Interconnection Facilities

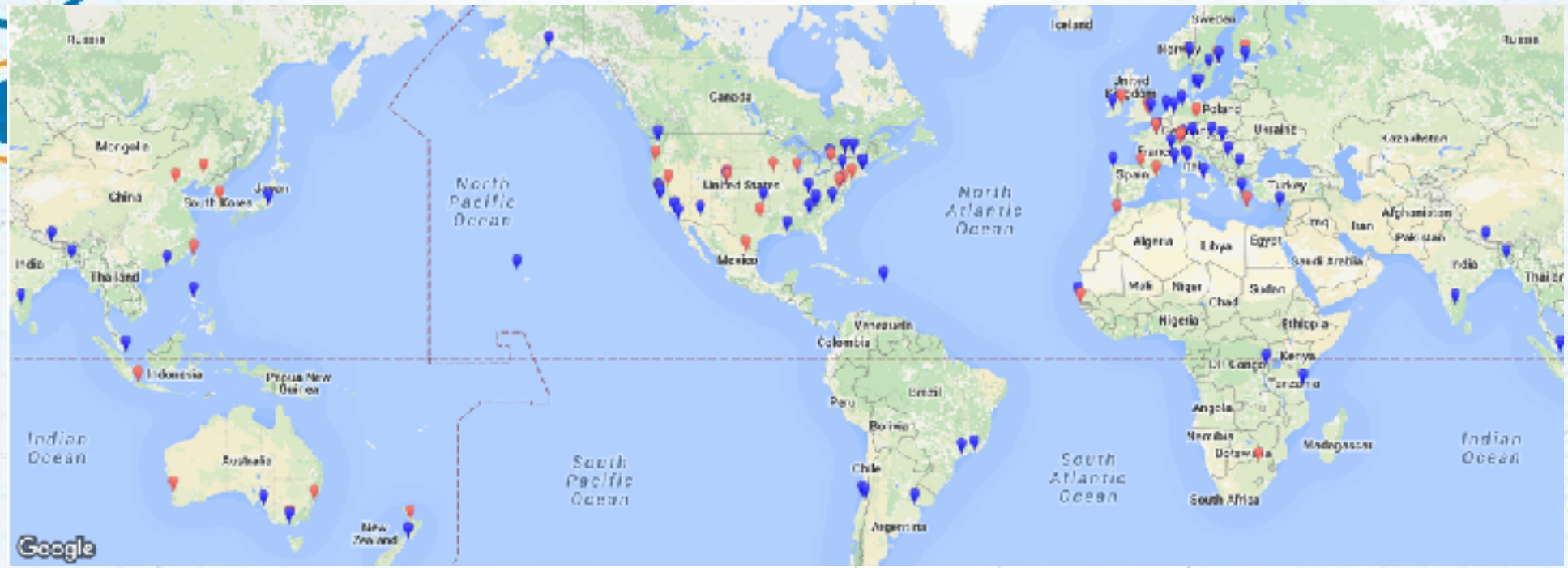
1. Assemble/maintain IXP database from published sources (PCH, PeeringDB)
2. Develop new techniques to infer engineering approach to interconnection
3. Alias resolution of interconnection IP addresses
4. Merge above techniques to generate a facility-aware map (CONEXT2015)
5. Open question: feasibility of scaling methods to hundreds of facilities



Providing Mapping-Related Data

- Publicly Available Mapping-Related Data (Most recent 2 years in DHS IMPACT)
  - IPv4 Routed /24 Topology, and associated DNS Names Dataset
  - IPv4 Prefix Probing Dataset (finer grained temporally)
  - Internet Topology Data Kits (ITDK)
  - IPv6 Topology and associated DNS Names Dataset
  - IPv4 Routed /24 AS Links (September 2007 - ongoing)
  - IPv6 AS Links (December 2008 - ongoing)
  - AS Relationships
  - AS Classification
  - AS to Organization

Archipelago Infrastructure



Legend: - Raspberry Pi - FreeBSD

- 181 monitors in 60 countries
  - 135 Raspberry Pi's
  - 83 have IPv6

Continent

67	North America
7	South America
56	Europe
22	Africa
20	Asia
9	Oceania

Organizations

57	academic
64	residential
33	commercial/business
27	network infrastructure
2	other

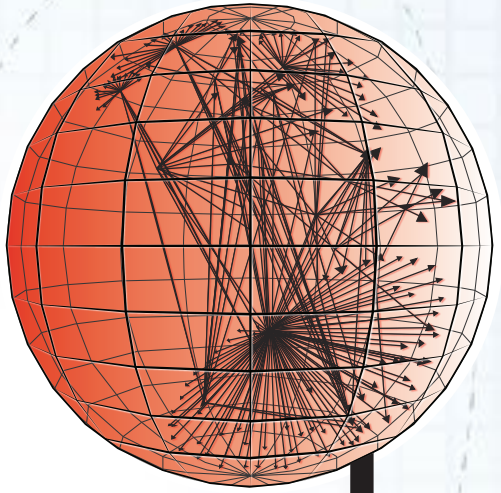
Benefits of Cybercartography

- Enhanced scientific understanding and technical capabilities for empirically grounded macroscopic assessment of the global Internet
- Comprehensive, trustworthy measurements of security-relevant properties and behavior of the global Internet
- Provides basis for data-focused services, products, tools and resources to advance the study of the Internet for a wide range of disciplines, led by today's imperative for improved cybersecurity
- Challenges looming ahead: effective data sharing, IoT, IPv6, CGN

How can you help?

- Use the data. Ask for more. Be specific.
- Use the measurement infrastructure platforms. Be scientific.
- Use the data integration platforms. Be situationally aware.
- Deploy monitoring infrastructure. Email info@caida.org
- Send promising students
- Teach promising students
- *Remember, you cannot secure what you cannot measure.*

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



caida

SDSC
SAN DIEGO SUPERCOMPUTER CENTER

UC San Diego

2017 Cyber Security R&D Showcase and Technical Workshop

July 11 - 13, 2017 | Washington, D.C.

Homeland Security

Science and Technology