Science of Internet Security: Technology and Experimental Research

k claffy / CAIDA/UCSD

SRI Arlington

12 April 2018

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), Ricky Mok, Paul Hick

Need

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

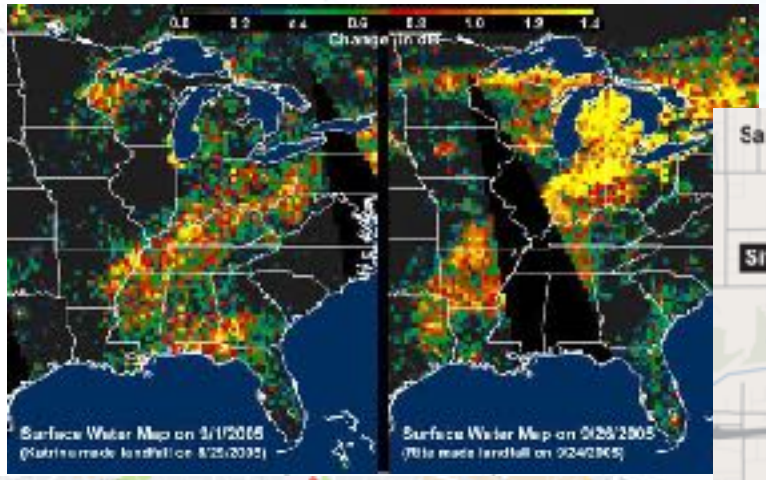
But we lack understanding of the structure, dynamics, and vulnerabilities of the global Internet.

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

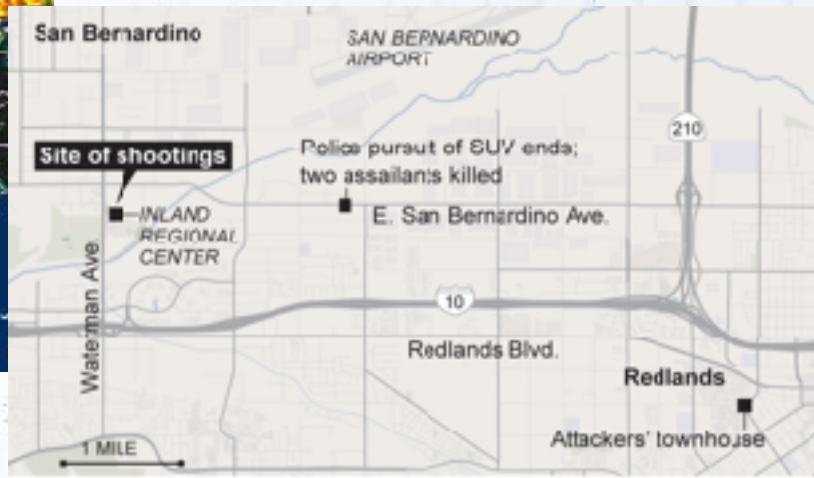
Real world crisis!

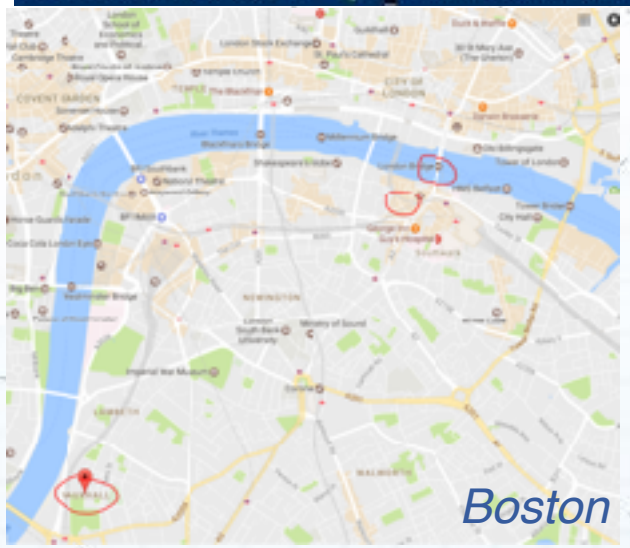
- **What do first responders do *first*?**
- **Find the crisis on a map!**
  - Where is it?
  - What is its scope?
  - What is its extent?
  - How fast is it moving?
  - How Do I Get There?

Katrina



San Bernardino, CA





Boston

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!

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

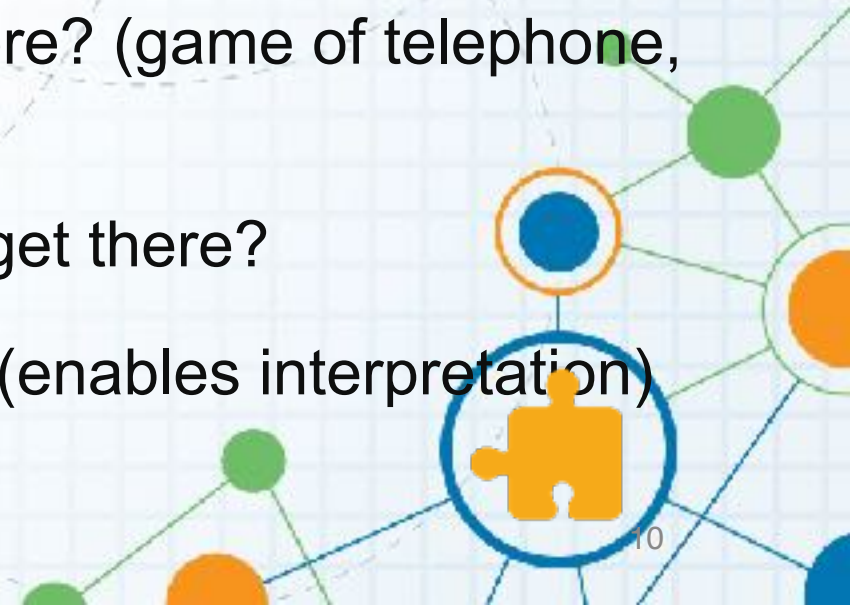
Guiding principle: ***The Internet was not designed to be measured.***

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

Archipelago Infrastructure

Legend: - Raspberry Pi - FreeBSD

- 211 monitors in 64 countries
  - 166 Raspberry Pi's
  - 98 have IPv6

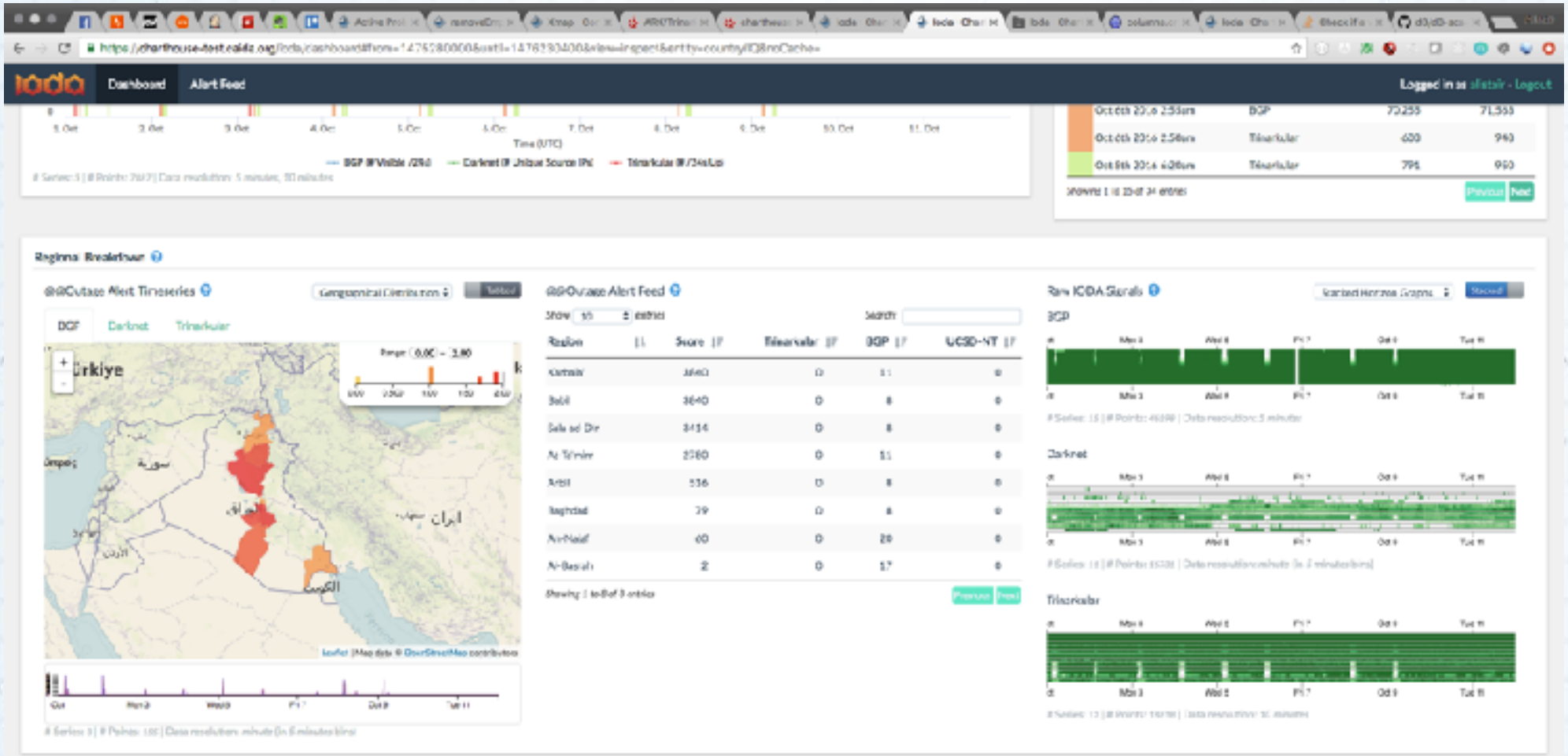
Continent

72	North America
9	South America
58	Europe
26	Africa
23	Asia
9	Oceania

Organizations

60	academic
81	residential
34	commercial/business
36	network infrastructure
0	other

Internet Outage Detection & Analysis

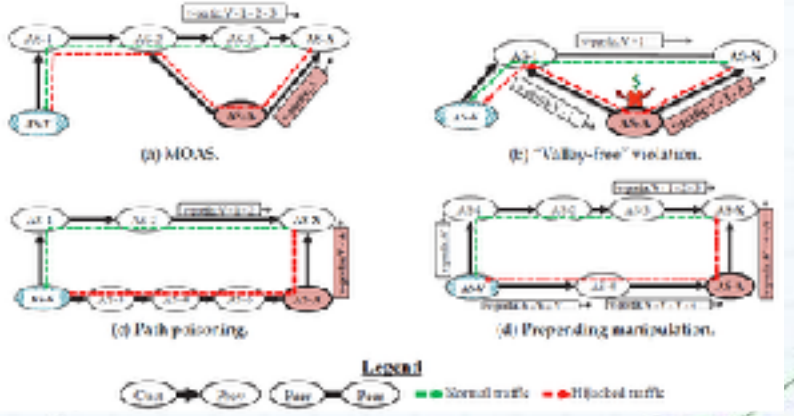


Screenshot of the IODA dashboard highlighting outages in Turkey

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

To measure outages, we need a baseline

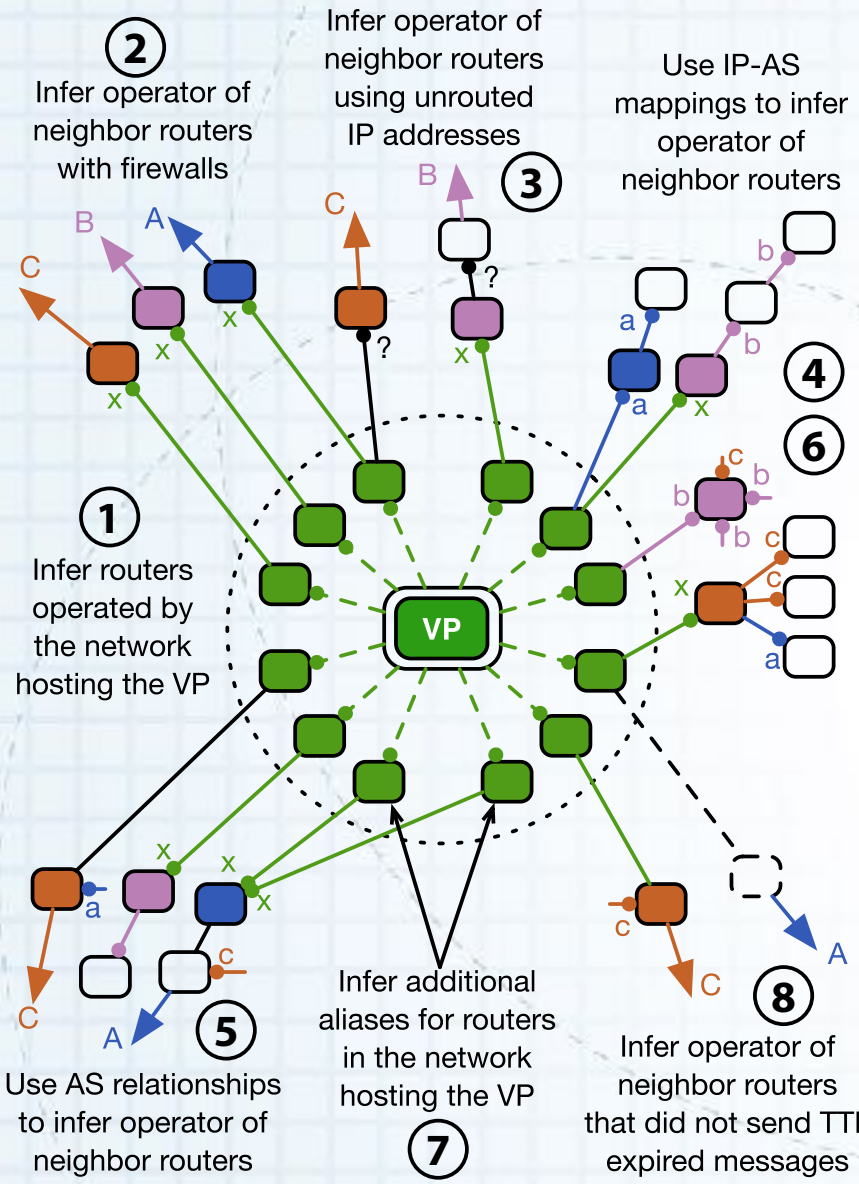
1. **Data acquisition and processing pipeline**
   a. sanitize BGP data from sliding 1-week of RV/RIS data (reconstruct routing tables of hundreds of operational routers worldwide with 5-min granularity)
   b. dynamically identify which IP addresses to probe (spatially and temporally fine-grained monitoring)
   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
   h. *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 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

Information about geographic locations of interconnection facilities, and autonomous systems (ASes) that have peering interconnections at those facilities.

Describe the methodology of mapping peering interconnections to interconnection facilities/IXPs

http://www.caida.org/data/as-facilities/

https://www.impactcybertrust.org/dataset_view?idDataset=832

Neutralizing BGP Hijacking

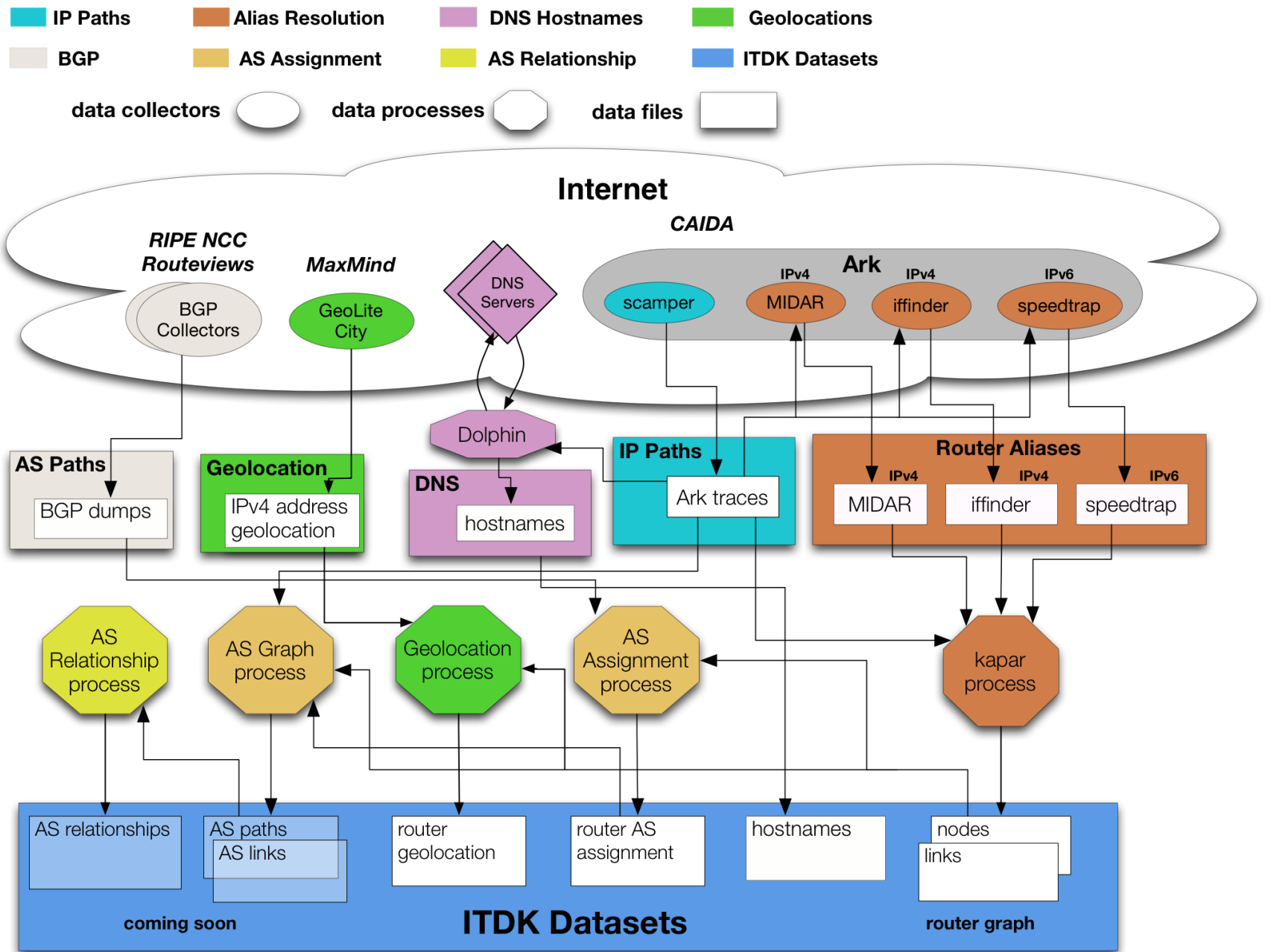
- Current approaches (e.g., RPKI, third-party services) lack:
  - comprehensiveness, allowing sophisticated attackers to evade detection,
  - accuracy, especially in the case of third-party detection,
  - real-time verification and mitigation of incidents, privacy in post-hijack response
- ARTEMIS: accurate and fast detection operated by the AS
  - Combines characteristics desirable to network operators such as comprehensiveness, accuracy, speed, privacy, and flexibility
  - Can neutralize prevalent classes of prefix hijacking within a minute
  - Open source (released soon), based on CAIDA's BGPStream
  - EU side sponsored by RIPE NCC

P. Sermpezis, V. Kotronis, P. Gigis, X. Dimitropoulos, D. Cicalese, A. King, and A. Dainotti, "ARTEMIS: Neutralizing BGP Hijacking within a Minute", Tech. rep., Center for Applied Internet Data Analysis (CAIDA), Jan 2018

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

ITDK: *Internet Topology Data Kit Process*



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

Competition

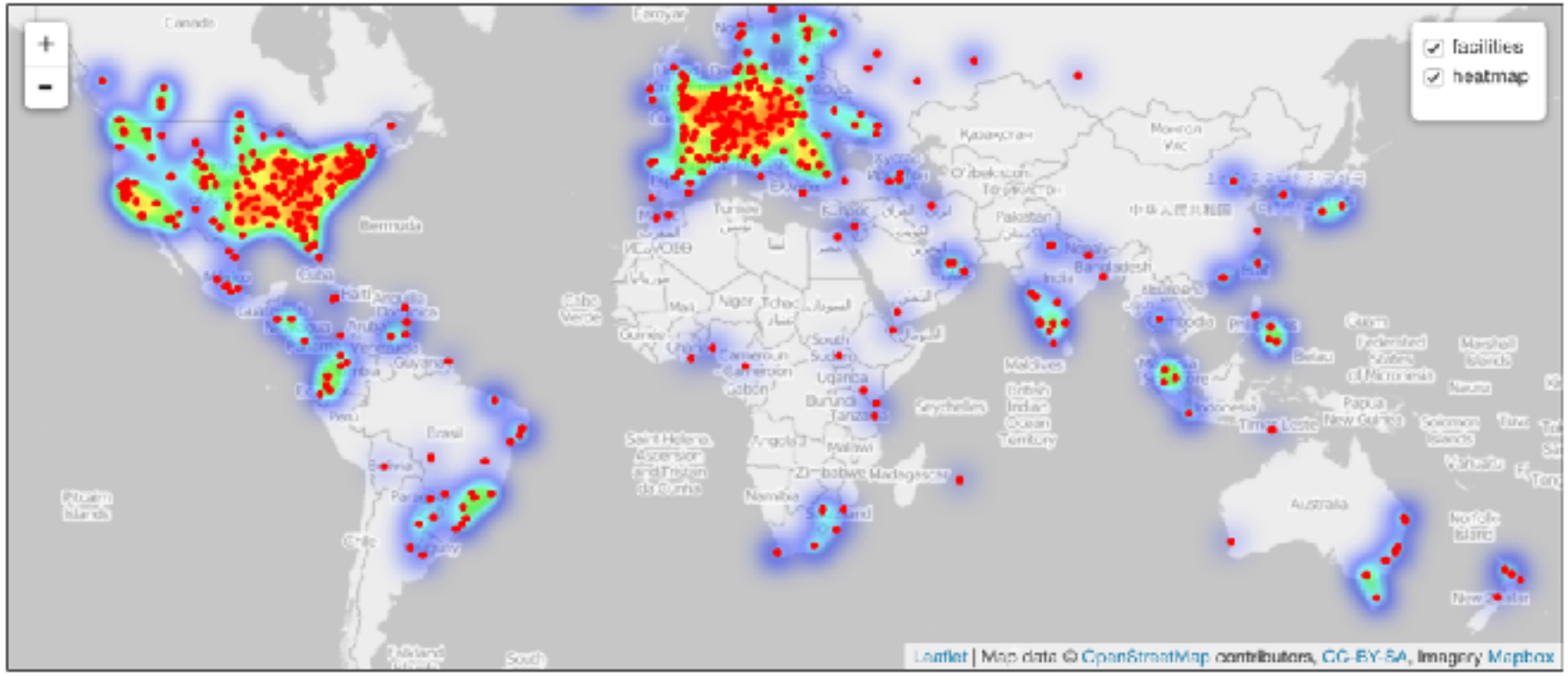
- Others infrastructures might be applied but stakeholders have different focus or lack community interest and access
  - RIPE Atlas (https://atlas.ripe.net)
  - Internet Atlas (http://internetatlas.org/)
  - ~~iPlane datasets (http://iplane.cs.washington.edu/data/data.html)~~
    - Data ends August 2016
  - zMap (https://zmap.io/), with select data (https://censys.io)
  - ISI Census (http://isi.edu/ant/address)
  - Dyn (http://www.dyn.com/) now Oracle

Current Status

- Period of performance ends May 23, 2018
- Task 1: Support for Macroscopic Security and Stability Monitoring and Analysis
  - Software/Data: IPv4 Prefix-Probing Traceroute Dataset
    http://www.caida.org/data/active/ipv4_prefix_probing_dataset.xml
  - Software: ARTEMIS
    http://www.caida.org/data/active/ipv4_prefix_probing_dataset.xml
- Task 2: Mapping Peering Interconnections at the Router Level
  - Software: updates to scamper
    http://www.caida.org/tools/measurement/scamper/
  - Data: CAIDA UCSD Border Mapping Dataset
    http://www.caida.org/data/active/bdrmap_dataset.xml

Current Status

- Task 3: Mapping Peering Interconnections at the Facility Level
  - Data: AS facilities Map
    http://www.caida.org/data/as-facilities/
- Data: IXP Dataset
  http://www.caida.org/data/ixps/



Current Status

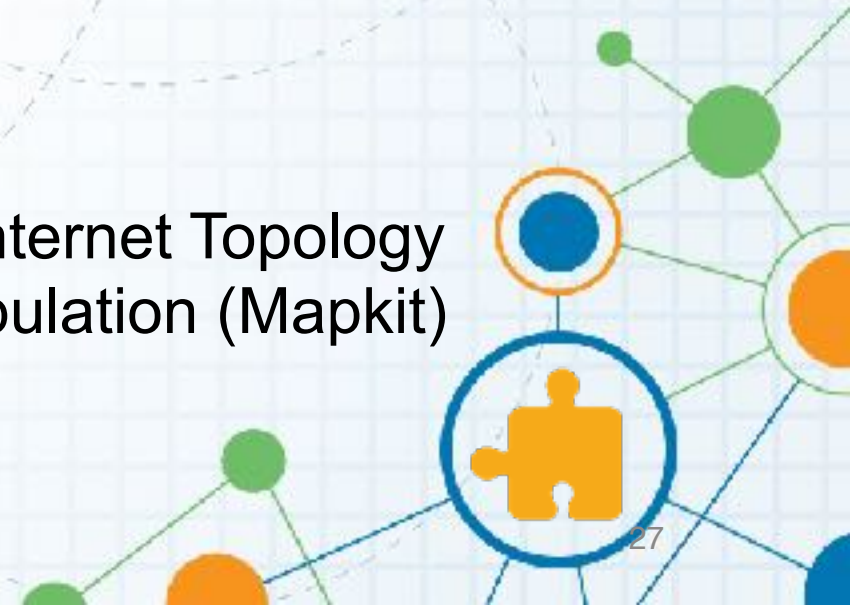
- Task 4: Measurements of TCP Behavior to Understand Security Vulnerabilities
  - Tech Report: Resilience of Deployed TCP to Blind FIN Attacks
    - http://www.caida.org/publications/papers/2017/resilience_deployed_tcp_blind/
- Task 5: Identifying Grey Market IPv4 Address Transfers
  - Paper: On IPv4 transfer markets: Analyzing reported transfers and inferring transfers in the wild
    http://www.caida.org/publications/papers/2017/ipv4_transfer_markets_wild/
  - Data: Complete Routed-Space DNS Lookups
    http://www.caida.org/data/active/complete_dns_lookups_dataset.xml

Current Status

- Task 6: Internet Router-Level Topology Mapping on Demand
  - Software: Improvements to MIDAR
    - http://www.caida.org/tools/measurement/midar/
  - Data/Query interface: Completion expected by end of contract

Transition/Completion Activities

- Experiments enabled by this work:
  - Internet Outage Detection and Analysis (IODA)
    http://ioda.caida.org/
  - ARTEMIS
    - plans to deploy at some Internet2 member sites
    - PI Dainotti attended I2 tech transition workshop and presented
  - NSF Interdomain Congestion Project
    http://www.caida.org/funding/nets-congestion/
  - NSF SATC Investigating the Susceptibility of the Internet Topology to Country-level Connectivity Disruption and Manipulation (Mapkit)
    http://www.caida.org/funding/satc-mapkit/



Publications

1. P. Sermpezis, V. Kotronis, P. Gigis, X. Dimitropoulos, D. Cicalese, A. King, and A. Dainotti, "ARTEMIS: Neutralizing BGP Hijacking within a Minute", Tech. rep., Center for Applied Internet Data Analysis (CAIDA), Jan 2018.

2. T. Holterbach, S. Vissicchio, A. Dainotti, and L. Vanbever, "SWIFT: Predictive Fast Reroute", in ACM SIGCOMM, Aug 2017.

3. M. Luckie and R. Beverly, "The Impact of Router Outages on the AS-level Internet", in ACM SIGCOMM, Aug 2017, pp. 488--501.

4. k. claffy and D. Clark, "The 9th Workshop on Active Internet Measurements (AIMS-9) Report", ACM SIGCOMM Computer Communication Review (CCR), vol. 47, no. 5, pp. 4, Oct 2017.

5. M. Jonker, A. King, J. Krupp, C. Rossow, A. Sperotto, and A. Dainotti, "Millions of Targets Under Attack: a Macroscopic Characterization of the DoS Ecosystem", in Internet Measurement Conference (IMC), Nov 2017.

6. M. Gharaibeh, A. Shah, B. Huffaker, H. Zhang, R. Ensafi, and C. Papadopoulos, "A Look at Router Geolocation in Public and Commercial Databases", in Internet Measurement Conference (IMC), Nov 2017.

Publications

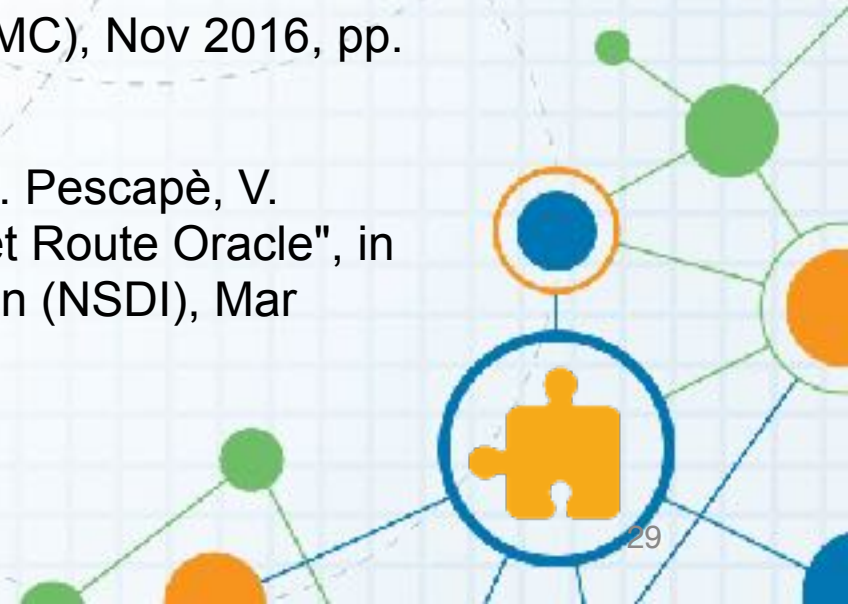
7. K. Levchenko, A. Dhamdhere, B. Huffaker, k. claffy, M. Allman, and V. Paxson, "PacketLab: A Universal Measurement Endpoint Interface", in Internet Measurement Conference (IMC), Nov 2017.

8. M. Luckie, "Resilience of Deployed TCP to Blind FIN Attacks", Tech. rep., Center for Applied Internet Data Analysis (CAIDA), Oct 2017.

9. V. Giotsas, G. Smaragdakis, C. Dietzel, P. Richter, A. Feldmann, and A. Berger, "Inferring BGP Blackholing Activity in the Internet", in Internet Measurement Conference (IMC), Nov 2017.

10. M. Luckie, A. Dhamdhere, B. Huffaker, D. Clark, and k. claffy, "bdrmap: Inference of Borders Between IP Networks", in Internet Measurement Conference (IMC), Nov 2016, pp. 381--396.

11. Í. Cunha, P. Marchetta, M. Calder, Y. Chiu, B. Schlinker, B. Machado, A. Pescapè, V. Giotsas, H. Madhyastha, and E. Katz-Bassett, "Sibyl: A Practical Internet Route Oracle", in USENIX Symposium on Networked Systems Design and Implementation (NSDI), Mar 2016.



Transition/Completion Activities

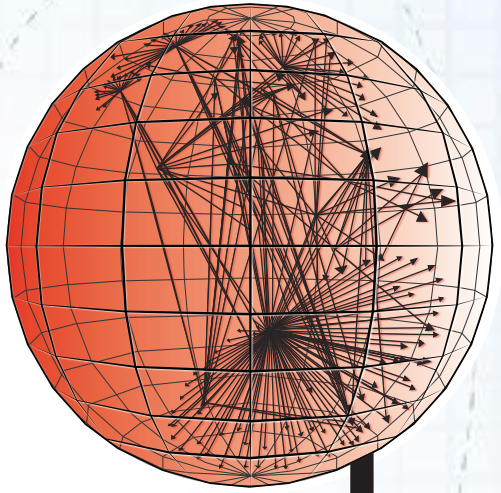
AIMS Workshop presentations representing technology/data use

(all slides at URL below):

- AS Rank
- Enabling machine-learning of router naming conventions
- Mapping AS borders on the Internet
- Mapping layer 2 (MPLS) topology
- Measuring interdomain congestion
- Alias resolution service for research community

http://www.caida.org/workshops/aims/1803/index.xml

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



caida

SDSC
SAN DIEGO SUPERCOMPUTER CENTER

UC San Diego