2019 S&T Cybersecurity and Innovation Showcase

Solutions Now I Innovations for the Future

Homeland Security
Science and Technology

Advancing Scientific Study of Internet Security and Topological Stability (ASSISTS DP)

kc claffy | CAIDA/UCSD

March 18, 2019

Homeland Security

Science and Technology

Funded Contract Information
This material is based on research sponsored by the Department of Homeland Security, Science and Technology Directorate via contract number FA8750-18-2-0049.

No Endorsement Notification
Any reference to any specific commercial products, processes, or services by trade name, trademark, manufacturer, or otherwise, does not constitute or imply its endorsement, recommendation, or favoring by the Department of Homeland Security or the United States Government.

Hyperlinked Web sites do not constitute endorsement by DHS of the Web site or the information, products, or services contained therein. DHS does not exercise any editorial control over materials on this website or the information on non-DHS Web sites.

Disclaimer Notification
The views, opinions, findings, conclusions, or recommendations expressed in this presentation are those of the authors and do not necessarily reflect the official policy or position of the Department of Homeland Security (DHS) or the United States Government. The publication of these views by DHS does not confer any individual rights or cause of action against the United States. Users of information in the materials assume all liability from such use.

Team Profile

- CAIDA undertakes research and operational activities to:
  - provide insights into Internet infrastructure, behavior, usage, and evolution,
  - build and maintain infrastructure and collaborative environments in which data can be acquired, analyzed, and (as appropriate) shared,
  - improve the integrity of the field of Internet science,
  - inform science, technology, and communications public policies.
- DHS enables and supports each of the above through IMPACT and other projects
- CAIDA's MPACT team: kc claffy and Alberto Dainotti (PIs), Daniel Andersen (System Admin), Paul Hick (Data Admin), Alistair King (Data Scientist), Alex Ma (Web Programmer), Marina Fomenkov (Project Manager)

SDSC
SAN DIEGO SUPERCOMPUTER CENTER

UC San Diego

Customer Need

Data to study security and stability-related events and identify cyber dependencies (TTA#1)

- Data collection, curation, hosting, stewardship, sharing
- New datasets supporting real-time incident identification
- Ongoing data collection to support studies of network structure and performance

Analysis system to identify, monitor, and mitigate infrastructure vulnerabilities (TTA#2)

- Web services and visual user interfaces for exploratory data analysis
- Software infrastructure for data storage, query, and transformation

Community outreach and service

- Disseminate information about cyber security data and analytics
- Host professional workshops and meetings
- User support

Need: What's Happening Now?

- Most data sharing happening in proprietary forms
- Limits (reproducible) science, inhibits academic or collaborative research on state of cybersecurity, handicaps operational security activities, policy efforts to improve trustworthiness of network
- Recent privacy legislation (GDPR) puts further constraints against data sharing (e.g., data about ownership of IP address or domain names), which handicaps law enforcement. Huge tussle ongoing.
- Even assessing risk of current situation requires data whose owners (even researchers) have counter-incentives to share

Approach (TTA#1) CAIDA as Data Provider

- Collecting and provisioning unique, critical data for the research community
- Internet Topology Data and Metadata (Ark Platform): 208 vantage points in 63 countries
  - 100 IPv6-enabled in 38 countries: 165 Raspberry Pis, 80 IPv6
  - Internet topology and performance, baselines and anomalies (congestion, hijacks)
- UCSD Network Telescope capturing background traffic (for 15 years)
  - As of Jan 2019, nearly 2TB of compressed traffic trace data per day.
  - Scanners, malware attacks, outages
  - 2015: 182 TB, 2016: 323 TB, 2017: 396 TB, 2018: 490 TB (27 Nov)
  - Approx. 3,846 days worth of data totaling 1,689 TB (1,536 TiB) archived at NERSC
- Passive Core Internet Traffic Traces (From Equinix-NYC)

Ark Platform: Topology Measurements

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

New: Data less than one year old available only through IMPACT.

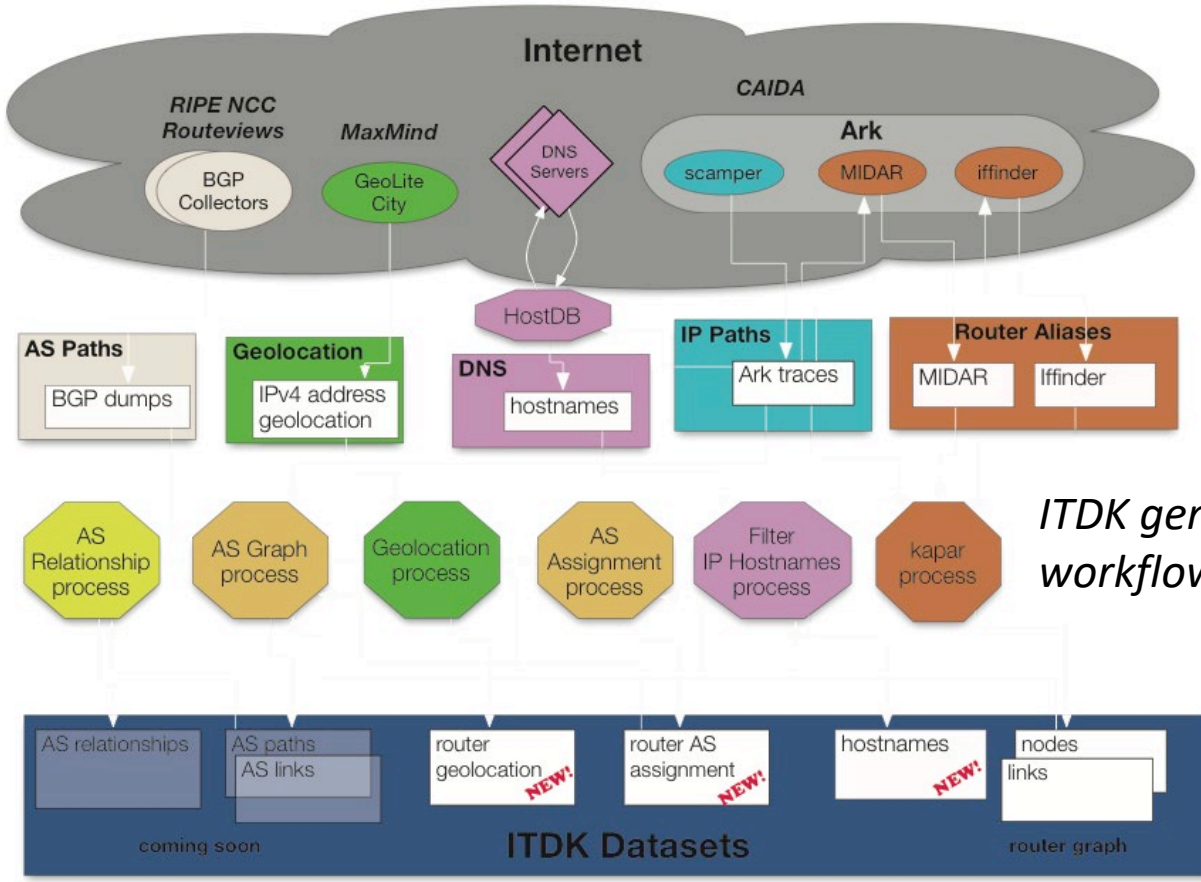
- 208 Ark monitors in 63 countries
- 100 IPv6-enabled
- 165 Raspberry Pis

Internet Topology Data Kits

http://www.caida.org/data/internet-topology-data-kit/

2017-08 & 2018-03, w/upgraded router-level mapping software.

- Traceroutes conducted on Ark and RIPE Atlas
- 2 related IPv4 + IPv6 router-level topologies
- Router-to-AS assignments
- Geographic location of each router
- DNS lookups of all observed IP addresses
- Supports baseline understanding of topology
- Longitudinal trend analysis
- Exploration of mapping algorithms and optimizations
- Educational use



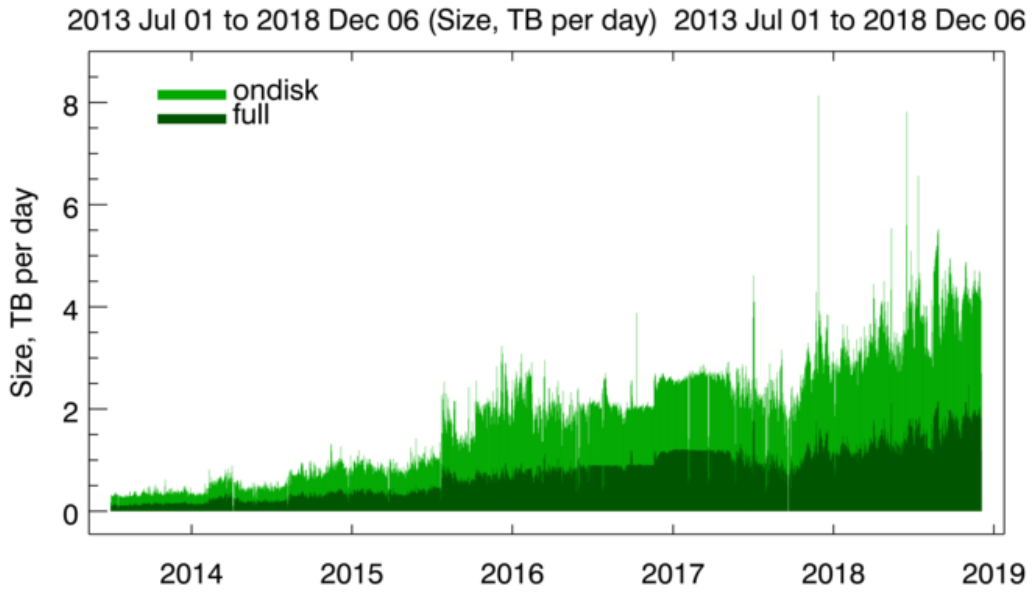
ITDK generation workflow

UCSD Real Time Telescope

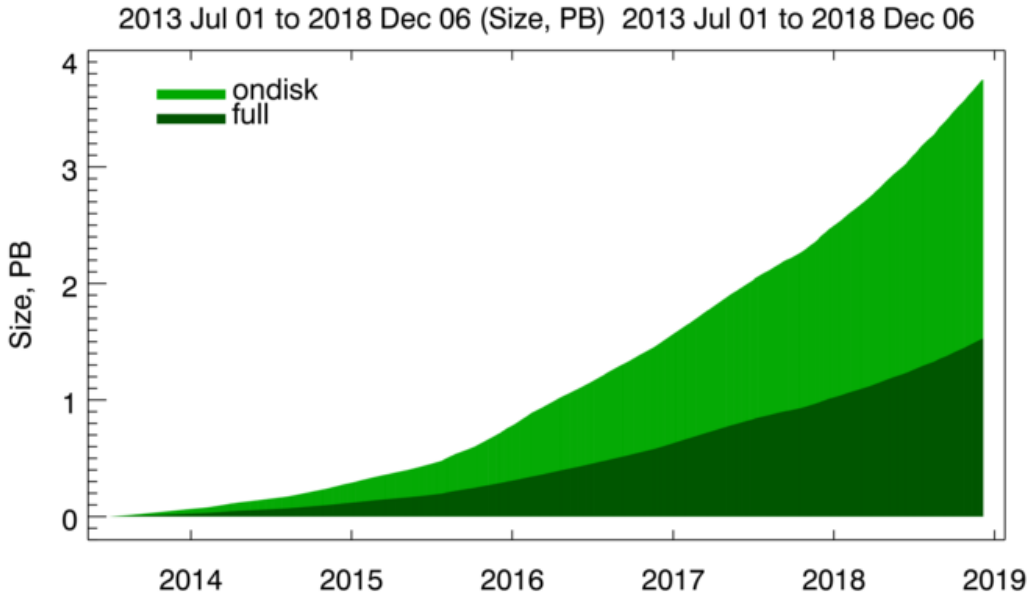
http://www.caida.org/data/passive/telescope-near-real-time_dataset.xml

- 491 TB added 1 Jan – 27 Nov 2018, Currently adding ~2TB/day
- 1.7 PB compressed data @ NERSC (since ‘08)
- Scanners, malware attacks, outages

Daily amount of data collected



Cumulative amount of data collected



Anonymized Passive Internet Traces

http://www.caida.org/data/passive/passive_dataset.xml

- Measurements from 2008 to Oct 2016.
- Starting March 2018, data contain anonymized traffic traces from CAIDA's new 10 Gb link Equinix-NYC monitor (Endance DAG cards)
- http://www.caida.org/data/monitors/passive-equinix-nyc.xml
- Raw traces stripped of payload, anonymized, split into 1-minute chunks
- Compressed size added March-April 2018: 582GB
- Collecting a single one-hour trace each month
- Used for traffic modeling and testing of security technologies

Other New Data Sets

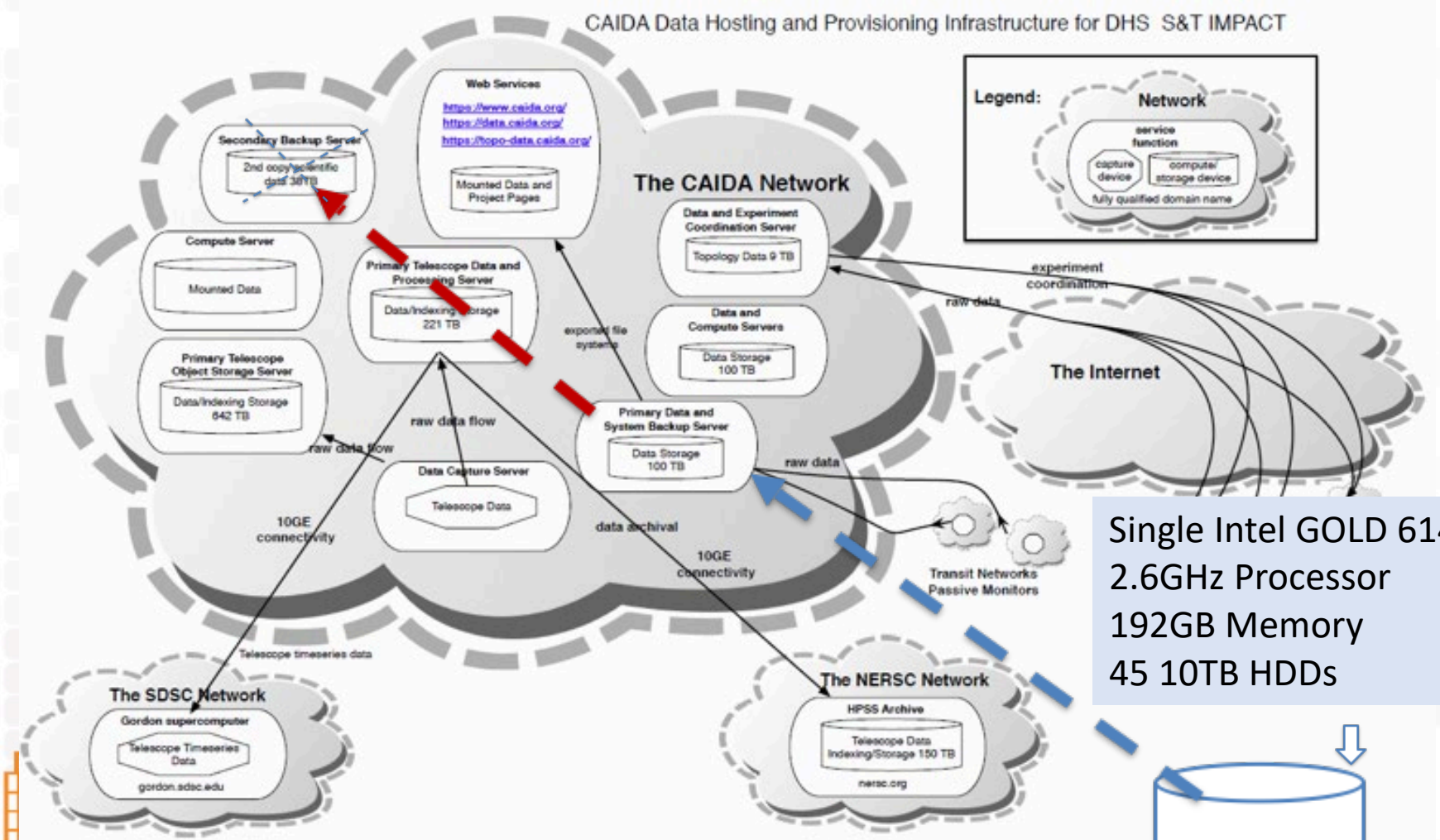
- 2018 IXP data set: info per 858 IXPs and facility, based on PeeringDB, HE, PCH DBs
- AS Facilities data: mapping of peering interconnections to facilities
- BGP community dictionary (2017)
- Metadata on DoS activity seen by UCSD Network Telescope
- All support situational awareness and infrastructure protection needs

All available via IMPACT portal, metadata also indexed at http://www.caida.org/data/

Numbers of targeted:
. attacks (black line)
. IP addresses (grey)
. /16 blocks (blue)
. ASNs (orange)

Data Hosting: Infrastructure

Installed and configured new storage server



Single Intel GOLD 6142, 16-core, 2.6GHz Processor
192GB Memory
45 10TB HDDs

Data Hosting: Size Added

From January-November 2018 we added 80K files (~1 PB)

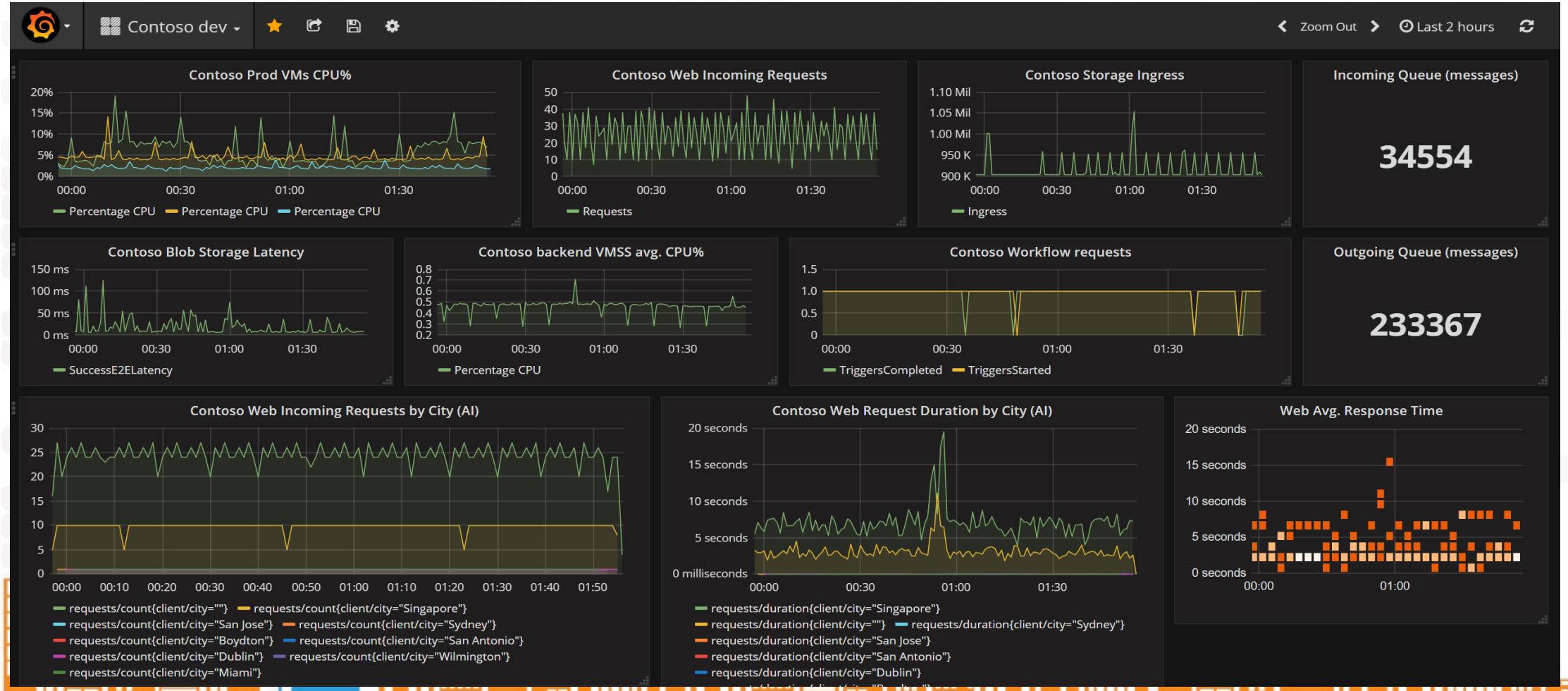
Data Collection	No. of files	On-disk size	Uncompressed size
Ark IPv4 Routed/24	60311	984GB	3.1 TB
Ark IPv4 Routed/24 DNS Names	241	18GB	66 GB
Ark IPv4 Prefix Probing	5844	632GB	2 TB
Ark Internet Topology Data Kits	25	7GB	56 GB
UCSD Network Telescope near real-time data	11630	377TB	1032 TB
Passive Traces	2258	1.9TB	4 TB
Cumulative totals	**80309**	**380 TB**	**1 PB**

New Tools Using IMPACT Data

- AS Rank – Autonomous Systems (AS) Rankings (RESTful API)
  https://as-rank.caida.org/
- Vela – on-demand topology measurement service and topology query system
  https://vela.caida.org/
  - Vela – web API – service for conducting measurements
    http://www.caida.org/projects/ark/vela/web-api/
    - Vela – MIDAR API – service for conducting alias resolution runs
      https://www.caida.org/projects/ark/vela/midar-api/
    - Vela - aliasq web-api - IP address alias query client
      https://www.caida.org/projects/ark/vela/aliasq-api/
- Spoofer - system to assess and report on the deployment of source address validation (SAV)
  https://spoofer.caida.org/
- **MANIC – Measurement and Analysis of Interdomain Congestion ** let's demo next ****
  https://manic.caida.org
- Internet Outage Detection and Analysis (IODA)
  https://ioda.caida.org/

MANIC platform

- Measurement and Analysis of Interdomain congestion
- Grafana front-end, API access to data also supposed



Competition/Alternatives

- Topology data: RIPE Atlas
  - Fantastic source of data, but cannot run experiments, do router-level mapping, study spoofing, interdomain congestion
- Telescope data: no other such project anymore (afawk)
- Passive traffic data from core backbone: nowhere else is such data available.

Lessons Learned

Significant infrastructure required to support security and stability research challenges of today's Internet.

IMPACT project demonstrates the effectiveness and amplifying power of responsible data sharing to support S&T research

Still a long way to go: Infrastructure necessary but also need platforms to correlate diverse data sets, and operational pilots to capitalize on them

Policy challenges: future privacy legislation is an opportunity to address data-sharing in support of cybersecurity in a more scalable, sustainable way

Contact Info

KC Claffy
CAIDA, University of California San Diego
kc@caida.org, alberto@caida.org
www.caida.org
@caidaorg

2019 S&T Cybersecurity and Innovation Showcase

Solutions Now I Innovations for the Future

Homeland Security
Science and Technology