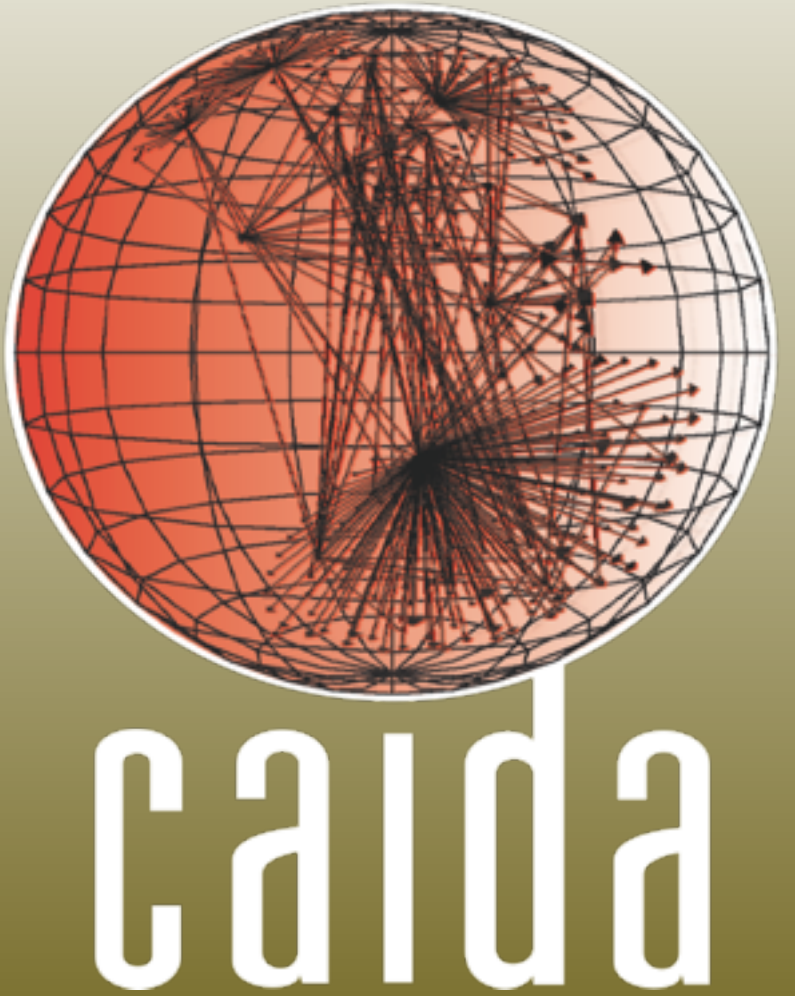
DHS PREDICT project: CAIDA update

Bradley Huffaker
PI k claffy, CAIDA
SRI
Washington D.C.
27 January 2015



DHS PREDICT project: CAIDA update

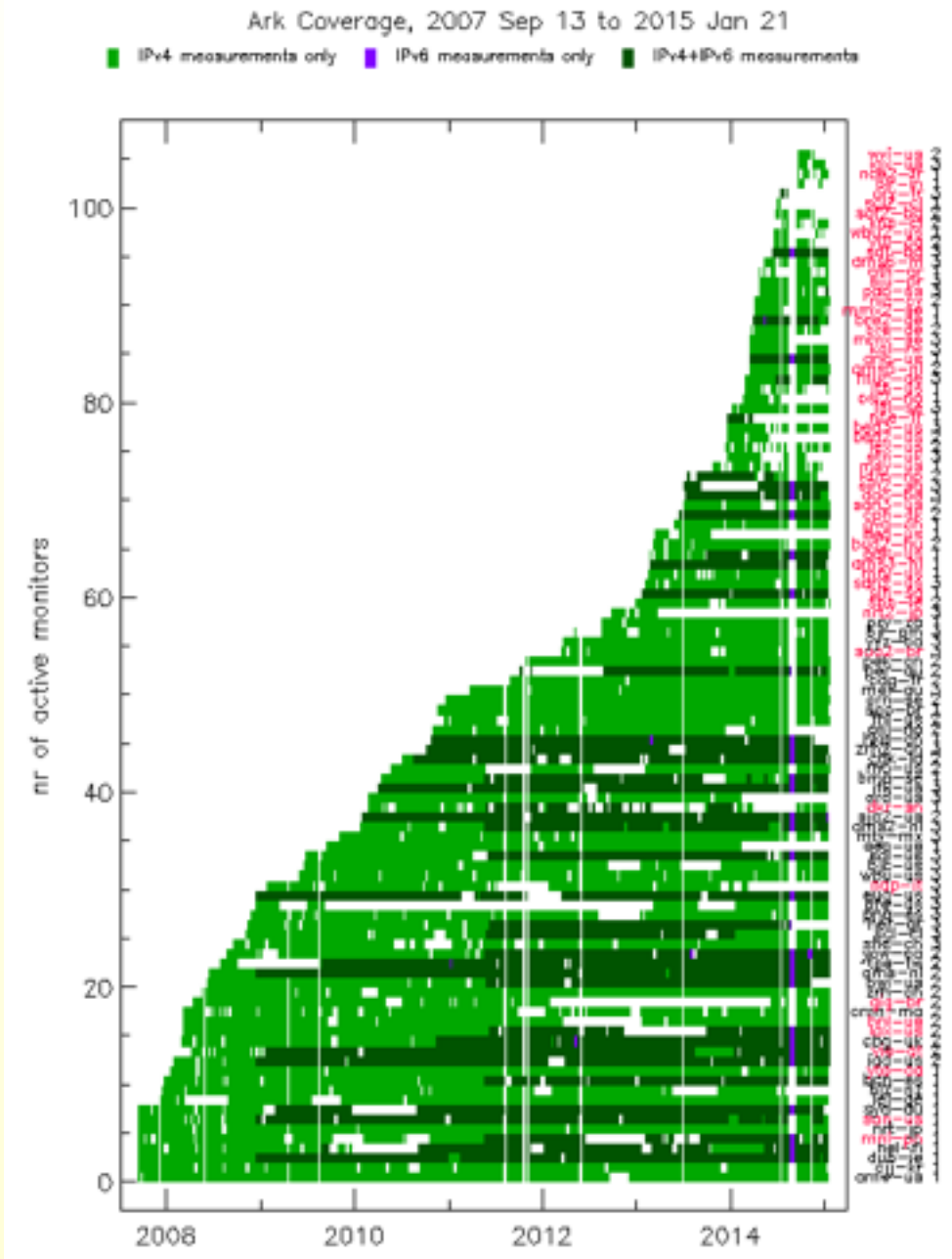
- Data collection activities
  - Ongoing measurements
  - Data storage status
  - Data dissemination statistics
- Other activities
  - New AUP for publicly downloadable data
  - AUP revisions
  - CREDS Workshop report
- Open issues
  - How new commercial attribute relates to our datasets
  - Losing tap on commercial link as it upgrades from 10Gig to 100 Gig

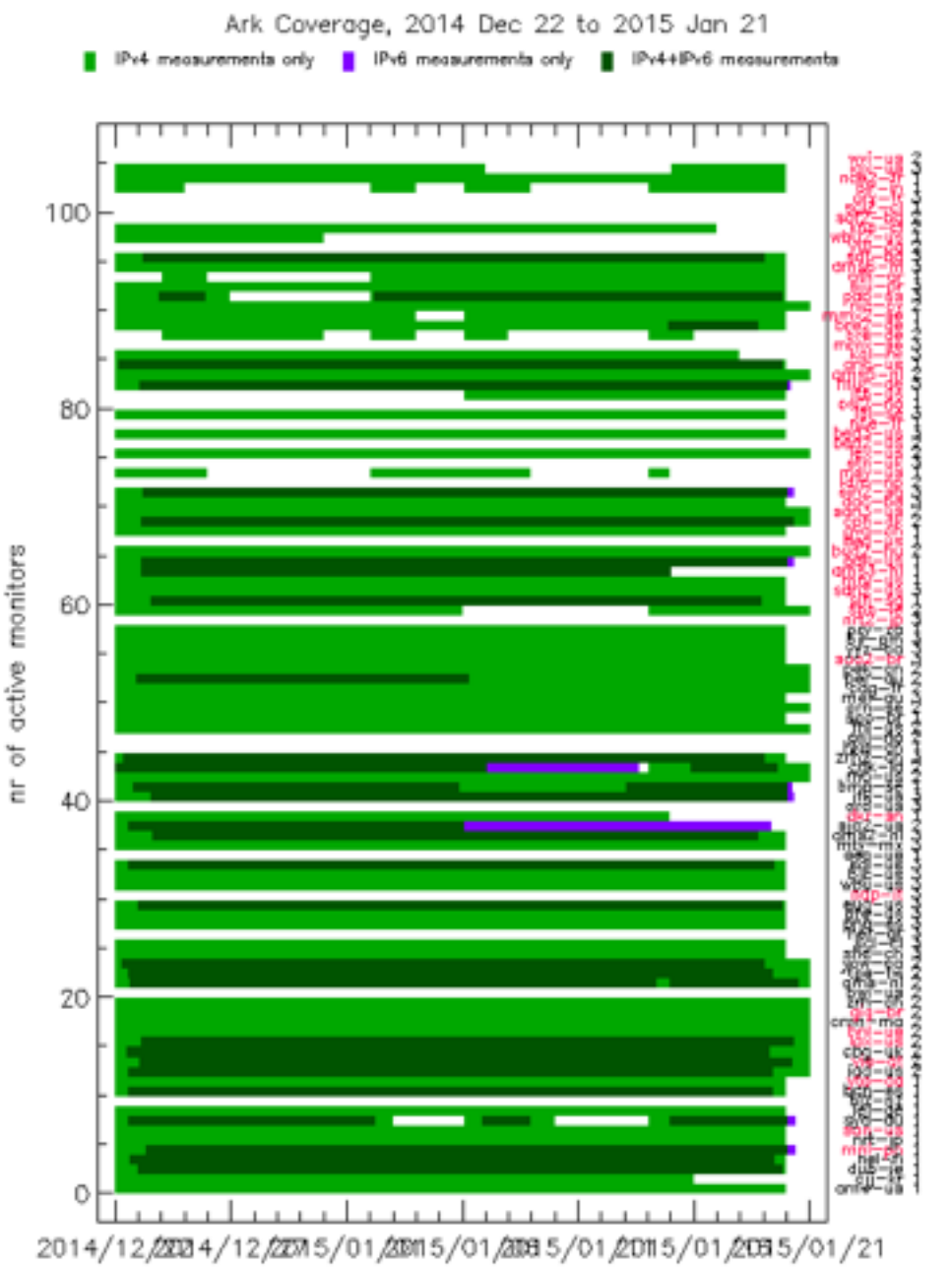
Ark Measurements

- Concurrent ongoing data collection
  - IPv4 and IPv6 topology
  - spoofer
  - continued, fine tuning congestion measurements
- Ark Platform (as of January 2015) - 106 monitors
  - 39 IPv6 enabled
  - 58 Raspberry PIs
- Derived data sets
  - ITDK (Dec 2014 released soon)
  - AS links, IPv4 and IPv6 daily
  - AS relationships
  - AS Ranking
  - AS to organization mapping

- Requests for restricted topology data
  - 2014: 65/46 requests received/approved
    - 227 unique users of the public (> 2yr) topology data
  - 2013: 162/118 prior to public data

Ark Data Coverage - Sept 2007 to Present





Ongoing Measurements: Internet Background Radiation

- UCSD Network Telescope
  - ~3-4 TB (May 2014)
  - 10 TB (Nov 2014)
  - 10 TB (Dec 2014)
  - 7.5 TB (so far in Jan 2015)
  - ~300 TB archived at NERSC
- Requests

  datasets
  - 2014: 36/54 requests approved/received
  - 2013: 29/44 requests approved/received

  near-real-time
  - 2014: 3/9 requests approved/received
  - 2013: 3/6 requests approved/received

Passive Trace Collection

- Passive infrastructure
  - two monitors with taps and Endace 10GE capture cards on two links at Equinix locations: San Jose and Chicago.
  - links upgrading to 100G: San Jose (Sep 2014), Chicago expected 2015Q4. **Current hardware cannot handle 100GB**
- Equinix traffic traces are worth continuing
  - 7 years of continuous packet samples (1 per month)
  - Available to researchers as "Anonymized Internet Traces"
  - Our **most popular restricted dataset**:
    - half of all requests for restricted CAIDA datasets
    - one third of all references to CAIDA datasets in literature
    - 80% of all data downloaded
    - usage examples: packet classification; testing of network devices; background traffic in IDS models.
  - Still **exploring options** for 100G capture.
    - Invea FPGA card (~$30k)
    - Endace Network Head end with 10G Endace card (~$100k)
    - Add $20k for storage

Public Datasets

- Added in 2014:

1. Ark IPv4 topology data older than 2 years
   http://data.caida.org/datasets/topology/ark/
2. All Ark IPv6 topology data
   http://data.caida.org/datasets/topology/ark/ipv6/
3. All historical skitter data
   http://data.caida.org/datasets/topology/skitter/
   http://data.caida.org/datasets/topology/skitter-itdk/
   http://data.caida.org/datasets/topology/skitter-aslinks/
4. AS-to-organization mapping dataset
   http://data.caida.org/datasets/as-organizations/

Data Storage

- Continue using NERSC for telescope data (no fee)
- Retired SDSC Cloud for backups, now local
- Continue to expand CAIDA storage

  Acquired in 2014:

  - 2U data server (loki)
    - 2x8 core CPUs w/ 256GB memory
    - 2x300GB mirrored root drives, and
    - 20x480GB SSDs for data

  - 4U disk shelf: (irori) 100TB
    - 24x4TB drives (with 21 empty bays for future expansion)
    - (2) 10 Gb network cards
    - replaces thoth (38 TB) as main data server, thoth now used to archive data replacing SDSC Cloud (25 TB)

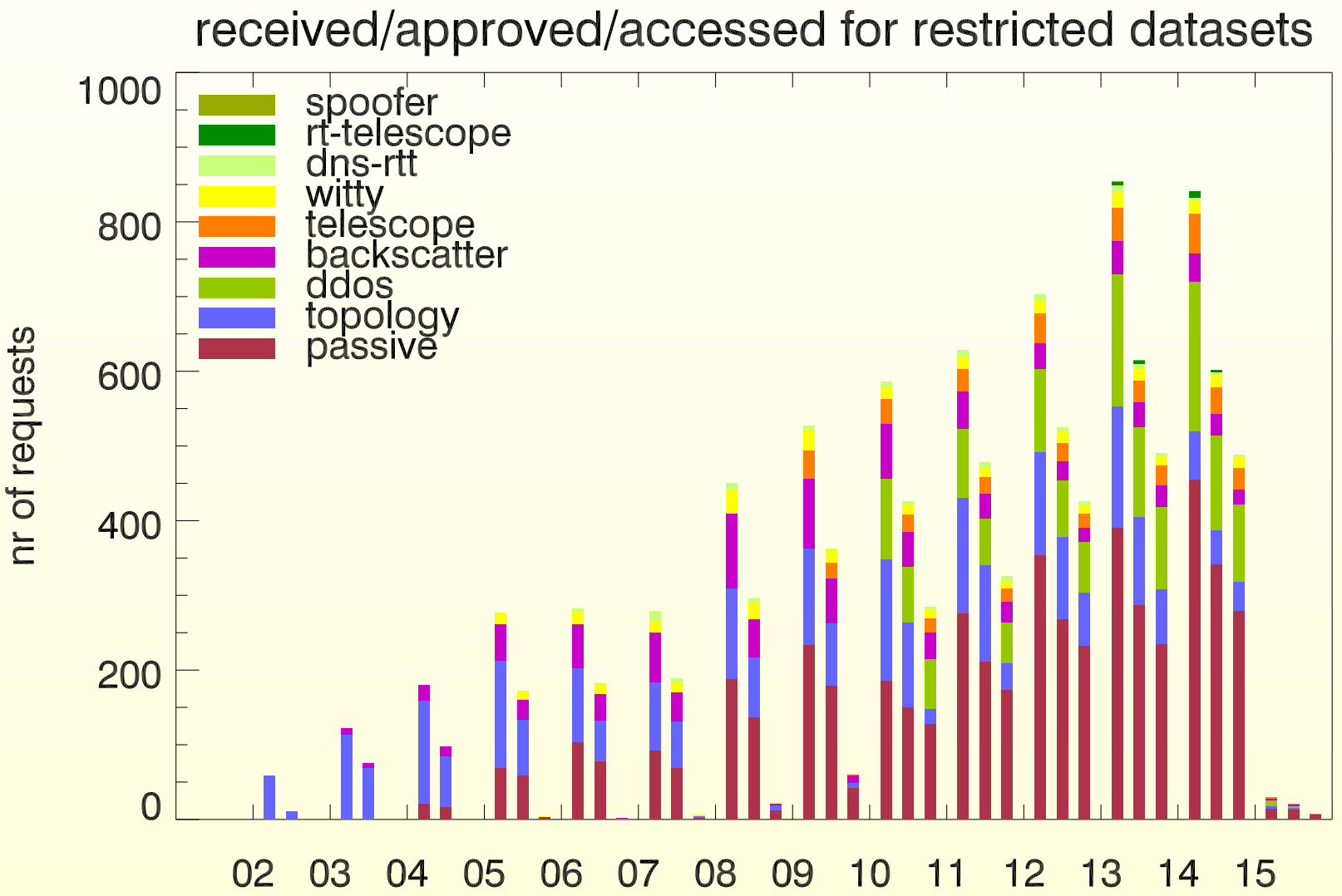
Data Access

- 2013 and 2014 Requests Summarized

	2013		2014	
Dataset	Received	Approved	Received	Approved
Topology	162	118	65	46
Passive	390	287	454	341
Telescope	44	29	54	36
Witty	21	15	17	17
Backscatter	45	33	38	28
Real-time telescope	6	3	9	3
Totals	668	485	637	471

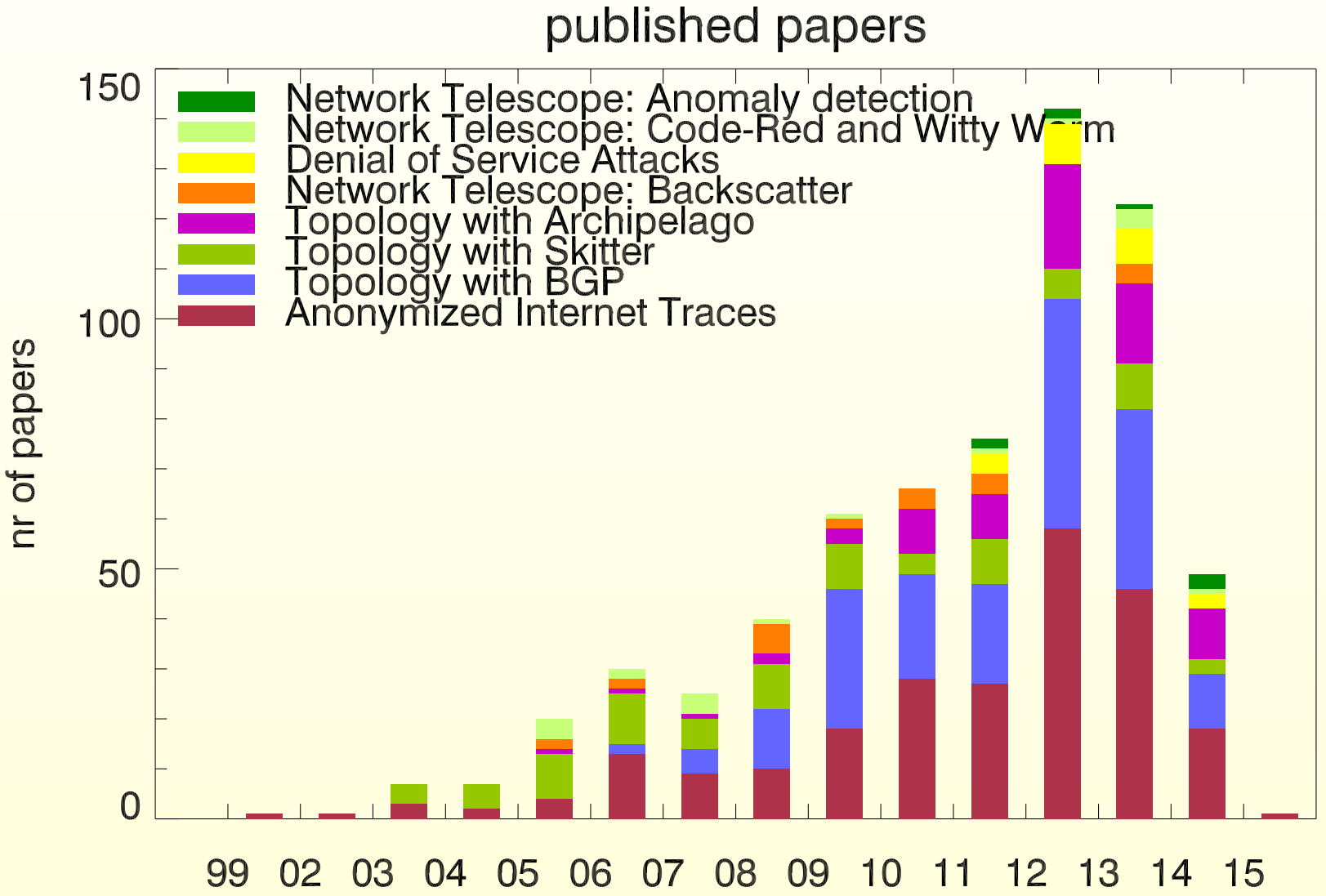
- Decrease in topology requests due to public availability of two years or older topology data.

Restricted Dataset Requests, 2014



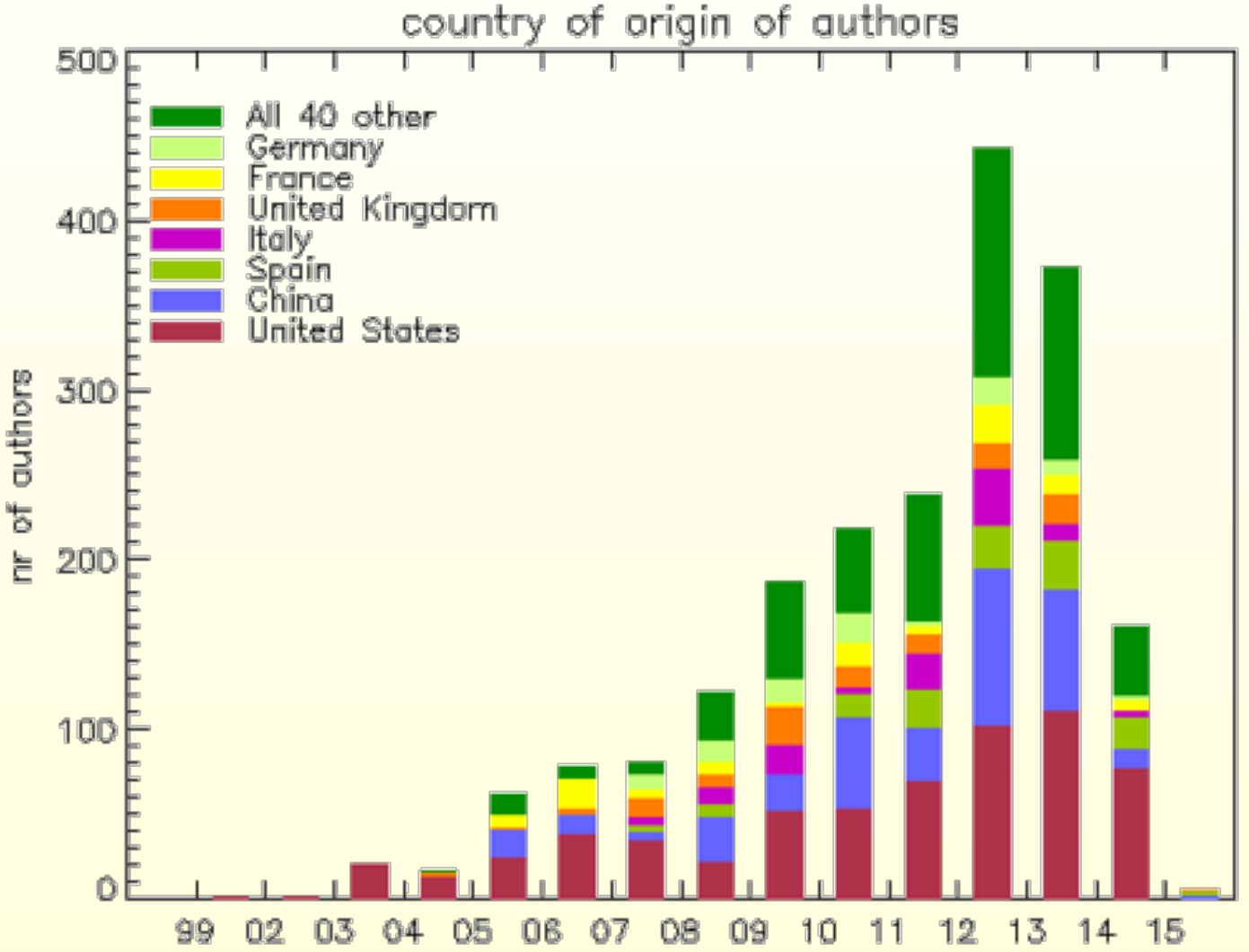
- requests for passive data are about half of all requests
- leveling off in 2013-2014 is due to making topology data public

non-CAIDA publications using PREDICT-related CAIDA data (that we know of)



- **2014 numbers are not final** as we have not completed the literature search for 2014 publications

non-CAIDA publications using PREDICT-related CAIDA data (that we know of)

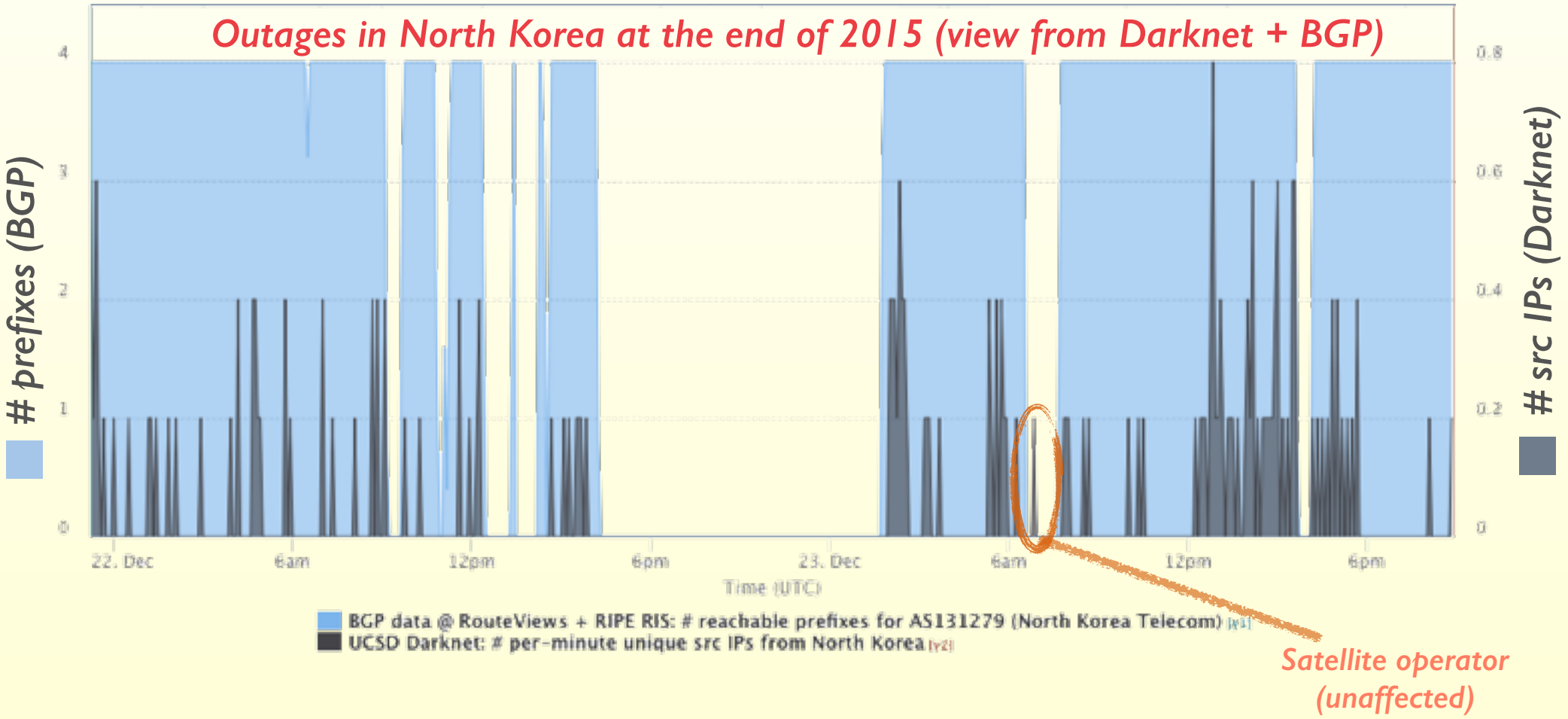


- **2014 numbers are not final** as we have not completed the literature search for 2014 publications

Internet Outages

IODA - Detection and Analysis of Internet Outages

NSF CNS-1228994 *(Sep 2012 - Aug 2015)*



E-RAID

Environment for Rapid Analysis of Internet Darknet traffic

code-to-data → *researcher-to-data&tools&expertise*

An integrated environment:

- ***multi-granular datasets***
- Charthouse: ***viz tools*** for agile inspection of post-processed data
- Corsaro: a modular data-processing pipeline allowing researchers to add their ***plugins***
- expert ***support*** to develop plugins and perform data analysis

Under submission to NSF CISE Research Infrastructure

Recent publications

- B. Huffaker, M. Fomenkov, and k. claffy, "**DRoP:DNS-based Router Positioning**", ACM SIGCOMM Computer Communication Review (CCR), vol. 44, no. 3, pp. 6--13, Jul 2014. [P,C4]
- M. Luckie, "**Spurious Routes in Public BGP Data**", ACM SIGCOMM Computer Communication Review (CCR), vol. 44, no. 3, pp. 15--21, Jul 2014. [C4]
- L. Zhang, A. Afanasyev, J. Burke, V. Jacobson, k. claffy, P. Crowley, C. Papadopoulos, L. Wang, and B. Zhang, "**Named Data Networking**", ACM SIGCOMM Computer Communication Review (CCR), vol. 44, no. 3, pp. 66--73, Jul 2014.
- In submission to Telecommunications Policy journal. D. Clark and k. claffy, "**Anchoring policy development around stable points: an approach to regulating the co-evolving ICT ecosystem**", Tech. rep., Massachusetts Institute of Technology, Aug 2014. [C4]
- D. Clark, S. Bauer, k. claffy, A. Dhamdhere, B. Huffaker, W. Lehr, and M. Luckie, "**Measurement and Analysis of Internet Interconnection and Congestion**", in Telecommunications Policy Research Conference (TPRC), Sep 2014. [C4]
- k. claffy and D. Clark, "**Platform Models for Sustainable Internet Regulation**", Journal of Information Policy, vol. 4, pp. 463--488, Sep 2014.

[P] used Predict Data
[C4] funded by C4

Recent publications

- Comments In the Matter of Protecting and Promoting the Open Internet, GN Docket No. 14-28. D. Clark, S. Bauer, and k. claffy, "**Approaches to transparency aimed at minimizing harm and maximizing investment**", Federal Communications Commission (FCC) Commission Documents, Sep 2014.
- A. Dainotti, K. Benson, A. King, k. claffy, E. Glatz, X. Dimitropoulos, P. Richter, A. Finamore, and A. Snoeren, "**Lost in Space: Improving Inference of IPv4 Address Space Utilization**", Tech. rep., Center for Applied Internet Data Analysis (CAIDA), Oct 2014. [P]
- k. claffy, D. Clark, and M. Wittie, "**The 6th Workshop on Active Internet Measurements (AIMS6) Report**", ACM SIGCOMM Computer Communication Review (CCR), vol. 44, no. 5, pp. 39--44, Oct 2014.
- V. Giotsas, M. Luckie, B. Huffaker, and k. claffy, "**Inferring Complex AS Relationships**", in Internet Measurement Conference (IMC), Nov 2014, pp. 23–30. [C4]
- M. Luckie, A. Dhamdhere, D. Clark, B. Huffaker, and k. claffy, "**Challenges in Inferring Internet Interdomain Congestion**", in Internet Measurement Conference (IMC), Nov 2014, pp. 15–22. [C4]

[P] used Predict Data
[C4] funded by C4

Recent publications

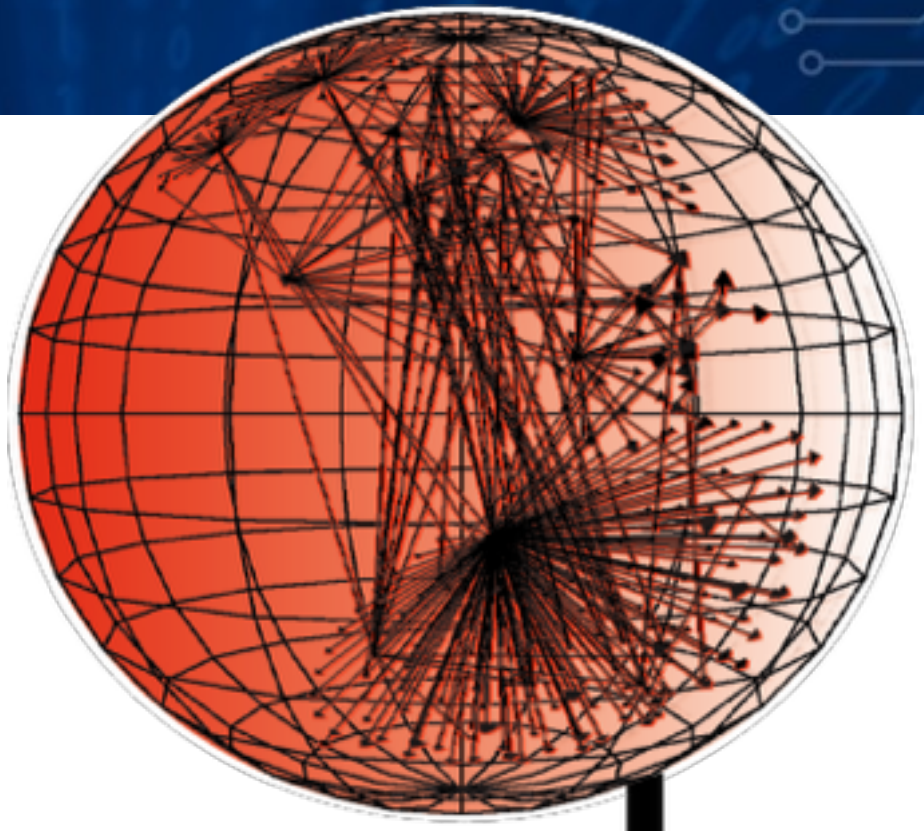
- L. Alt, R. Beverly, and A. Dainotti, "**Uncovering Network Tarpits with Degreaser**", in Annual Computer Security Applications Conference (ACSAC), Dec 2014.
- Originally in-proceedings of the Internet Measurement Conference (IMC) in 2011, this paper is "in press" for IEEE/ACM Transactions on Networking (ToN), and was published as an early access article in December 2013. A. Dainotti, C. Squarcella, E. Aben, K. Claffy, M. Chiesa, M. Russo, and A. Pescapè, "**Analysis of Country-wide Internet Outages Caused by Censorship**", IEEE/ACM Transactions on Networking, 2014. [P]
- Originally in-proceedings of the Internet Measurement Conference (IMC) in 2012, this paper is "in press" for IEEE/ACM Transactions on Networking (ToN), with an Early Access Article published in January 2014. A. Dainotti, A. King, K. Claffy, F. Papale, and A. Pescapè, "**Analysis of a "/0" Stealth Scan from a Botnet**", IEEE/ACM Transactions on Networking, 2014. [P]
- k. claffy, J. Polterock, A. Afanasyev, J. Burke, and L. Zhang, "**The First Named Data Networking Community Meeting (NDNcomm)**", Tech. rep., Named Data Networking (NDN), Jan 2015.

[P] used Predict Data
[C4] funded by C4

Contact Information

Bradley Huffaker
bradley@caida.org
PI: k claffy
kc@caida.org

http://www.caida.org/



caida
www.caida.org