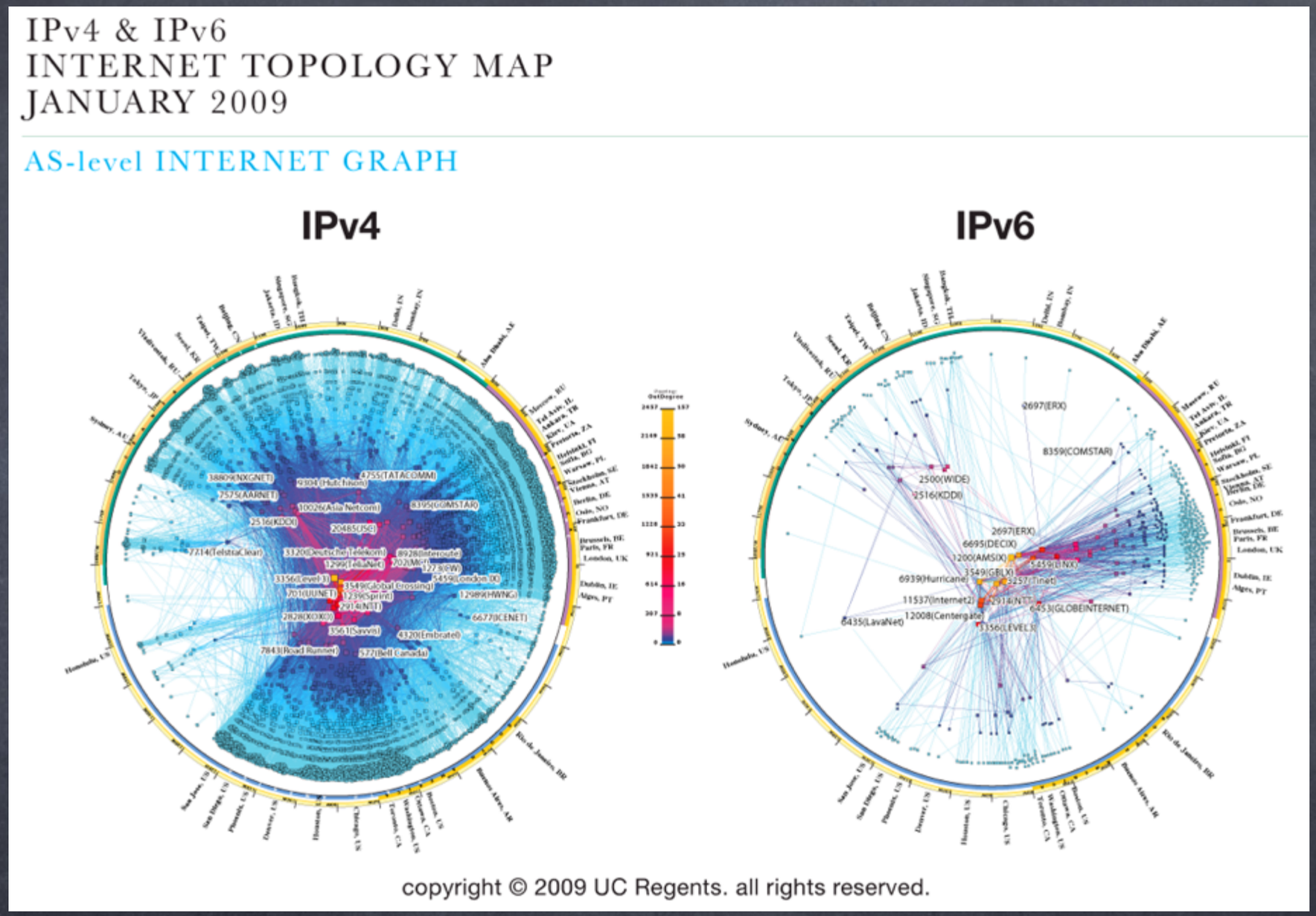
CAIDA Updates



kc claffy

CAIDA
UCSD CSE sysnet
3 March 2010

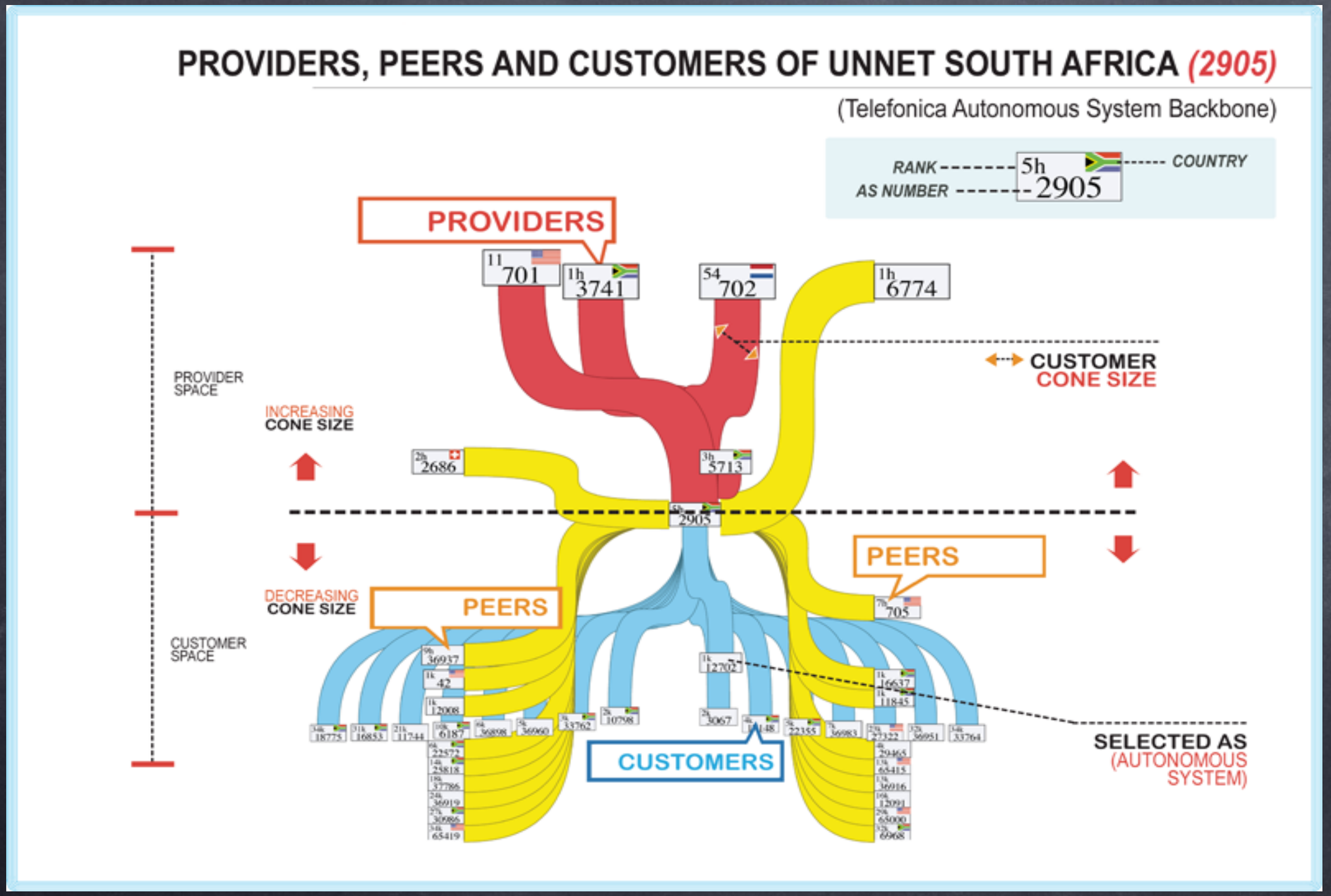
Overview

- CAIDA:
  - conducts research
  - informs policy
  - builds infrastructure
  - develops tools
  - collects and curates data
  - conducts workshops
  - has potential student projects

Research

- Macroscopic Topology Project
  - Archipelago active measurement infrastructure
    - IPv4 and IPv6 topology discovery
      - DNS annotation
      - AS-level, router-level, dual graphs
- Alias Resolution
- Router to AS assignment

Research

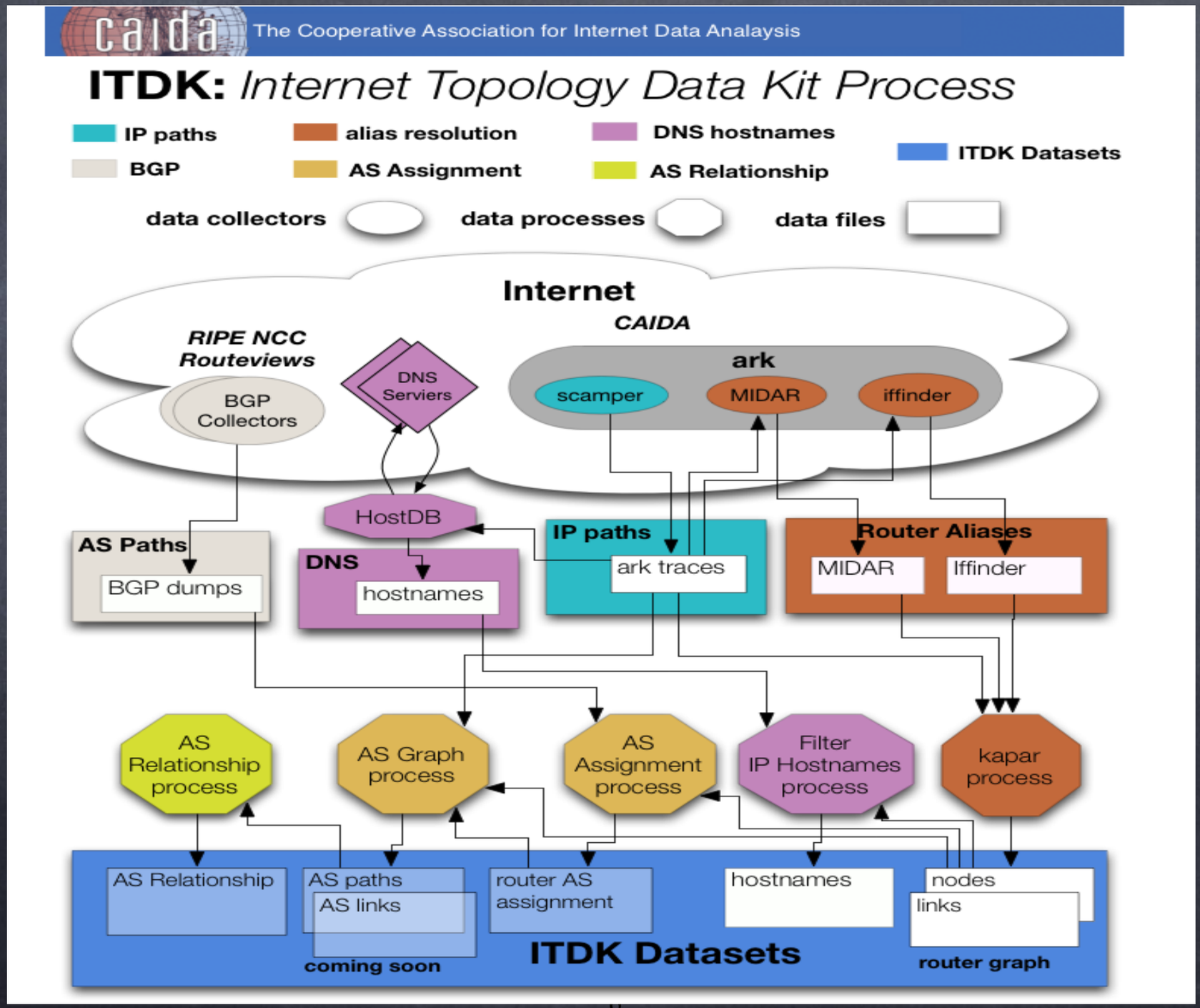

PROVIDERS, PEERS AND CUSTOMERS OF UNNET SOUTH AFRICA (2905)
(Telefonica Autonomous System Backbone)
RANK
AS NUMBER
5h
2905
COUNTRY
PROVIDERS
11 701
1h 3741
54 702
1h 6774
PROVIDER SPACE
INCREASING CONE SIZE
CUSTOMER CONE SIZE
2h 2686
3h 5713
5h 2905
PEERS
7h 705
DECREASING CONE SIZE
PEERS
CUSTOMER SPACE
CUSTOMERS
SELECTED AS (AUTONOMOUS SYSTEM)

Research (cont)

- Router to AS mapping
- Alias Resolution
  - Collapse interfaces to produce router-level graph
    - APAR-> kapar
    - RadarGun->MIDAR
- MAARS – regularly updated AS-level, router-level and dual graph – ITDKs
- Heuristics to assign routers to ASes creating dual graph

Research (cont)

- Internet Topology Data Kit (ITDK) Process



Research (cont)

- Routing on Complex Networks
  - Navigability
- M. Boguna, D. Krioukov, kc claffy, "Navigability of complex networks", Nature Physics, vol 5, no. 1, pp-74-80, January 2009.
  http://www.caida.org/publications/papers/2009/navigability_complex_networks/
- M. Boguna, D. Krioukov, "Navigating ultrasmall worlds in ultrashort time", in Physical Review Letters, vol 102, no 058701, 2009.
  http://www.caida.org/publications/papers/2009/navigating_ultrasmall/
    -

Research (cont)

- Routing on Complex Networks
  - Curvature and Temperature
- D. Krioukov, F. Papadapoulos, A. Vahdat, M. Boguna, "Curvature and temperature of complex networks", Physical Review E, v80, 035101(R), 2009.
  http://www.caida.org/publications/papers/2009/curv_temp_complex_nets/
  - Greedy forwarding
- F. Papadapoulus, D. Krioukov, M. Boguna, A. Vahdat, "Greedy forwarding in scale free networks embedded in hyperbolic metric spaces", in ACM SIGMETRICS Performance Evaluation Review, vol. 37, no. 2, pp. 15-17, Oct 2009.
  http://www.caida.org/publications/papers/2009/greedy_forwarding_embedded/

Research (cont)

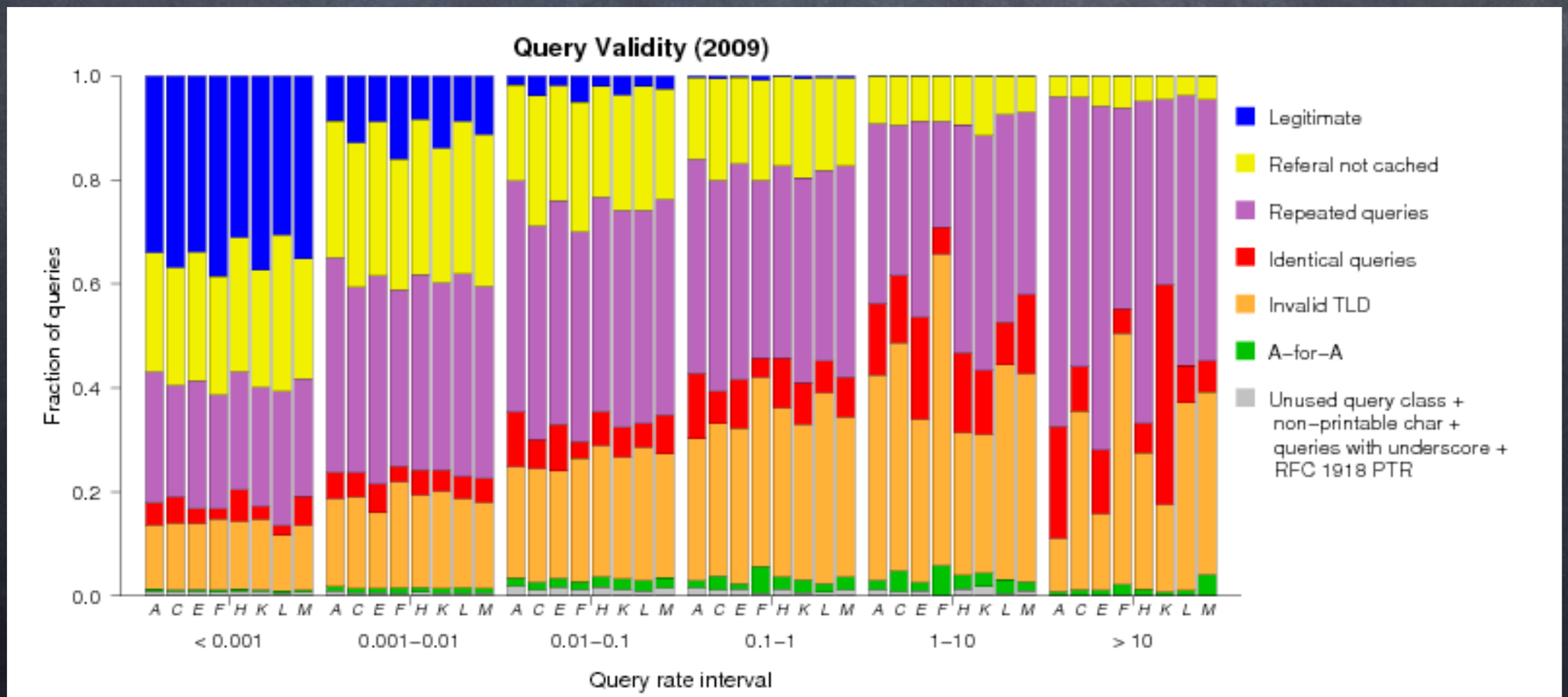
Routing on Complex Networks

- M. Kitsak, L. Gallos, S. Havlin, F. Liljeros, L. Muchnik, H.E. Stanley, H. Makse, "Identifying influential spreaders in complex networks", posted to the Physics and Society section of arxiv.org in January 2010.

http://www.caida.org/publications/papers/2010/influential_spreaders/

Research (cont)

- Domain Name System (DNS) S. Castro, M. Zhang, W. John, D. Wessels, kc claffy, "Understanding and preparing for DNS evolution", to be presented at the 2nd International Traffic Monitoring and Analysis (TMA'10) Workshop colocated with PAM2010.
  http://www.caida.org/publications/papers/2010/understanding_dns_evolution/



Research (cont)

- Internet Traffic Analysis
  - Traffic classification overview
    http://www.caida.org/research/traffic-analysis/classification-overview/
- Routing Asymmetry Study
  http://www.caida.org/research/traffic-analysis/asymmetry/

Policy

- kc claffy, "Historical and Architectural Context for Traffic Management Needs Today", presented at the FCC Technical Advisory Process workshop on December 8, 2009.|
  http://www.caida.org/publications/presentations/2009/traffi c_historical_context/

Policy (cont)

- E. Kenneally, kc claffy, "*An Internet Data Sharing Framework For Balancing Privacy and Utility*", in the proceedings of *Engaging Data: First International Forum on the Application and Management of Personal Electronic Information*, MIT, Oct. 12-13, 2009.

  http://www.caida.org/publications/papers/2009/engaging_data/

- E. Kenneally, M. Bailey, D. Maughan, "*A Framework for Understanding and Applying Ethical Principles in Network and Security Research*", in the proceedings of the *Workshop on Ethics in Computer Security Research (WECSR 2010)* in January 2010

  http://www.caida.org/publications/papers/2010/framework_ethical_research/

Policy (cont)

- E. Kenneally and kc claffy, "*Dialing privacy and utility: a proposed data-sharing framework to advance Internet research*", in submission to *IEEE Security & Privacy* special issue, July 2010.

http://www.caida.org/publications/papers/2009/dialing_privacy_utility/

PS2/Privacy Risk	Public Disclosure	Compelled Disclosure	Malicious Disclosure	Government Disclosure	Misuse	Inference Risk	Re-ID Risk
Authorization		X	X		X	X	X
Transparency	X	X	X	X	X		
Law Compliance			X			X	X
Access Limitation		X			X	X	X
Use Specification		X	X		X	X	
Minimization							X
Audit Tools	X	X	X	X	X	X	X
Redress	X	X	X	X	X	X	X
Oversight		X	X			X	X
Data Quality	X	X	X	X			X
Security		X				X	X
Training/Education		X	X			X	X
Impact Assessment	X	X	X	X	X		

Minimiz.Tech.	Is Purpose Worthwhile?	Is there a need?	Is it already being done?	Are there alternatives?	Is there a scientific basis?	Can results be acted upon?	Can DS & DP implement?	Reasonable education costs?	Forward & backward controls?	No new privacy risks created?	No free rider problem created?
Not Sharing	X	X	X	X	X	X	X				
Delete All	X	X	X	X	X	X	X		X		
Delete Part	X	X		X	X		X		X	X	
Anonymize	X	X	X	X	X		X	X	X	X	
Aggregate	X	X	X	X	X				X	X	
Mediate (SC2D)	X						X	X			X
Age Data	X	X	X	X	X		X			X	
Limit Quantity	X	X	X	X	X	X	X		X	X	
Layer Anonymization	X	X	X		X	X	X	X	X		

Policy: IRB (re)review

- We submitted our 2nd IRB application to UCSD in November 2009, provided requested clarifications in December and were told in January 2010 that our protocol does not qualify as human subjects research.

The Committee at the January 6, 2010 Institutional Review Board meeting reviewed your December 14, 2009 correspondence. The Committee acknowledged that the PI had provided a thoughtful and substantive response to their letter dated December 2, 2009; however, the Committee voted unanimously that the proposed activity does not satisfy the regulatory definition of human subjects research and therefore falls outside the jurisdiction of the IRB. For further clarification we offer the following:

It is the Institutional Review Board's charge to review research that presents a testable hypothesis and study design that will add to generalizable knowledge and takes into account protection of living individuals under appropriate informed consent. Based on the information submitted, the proposed activity is not actively seeking information from human subjects, but rather computer generated information. Should there be a decision in the future to develop, for research purposes, a hypothesis and other research related design methods that utilizes information gathers from individuals, the Committee recommended that a new application be submitted for IRB review.

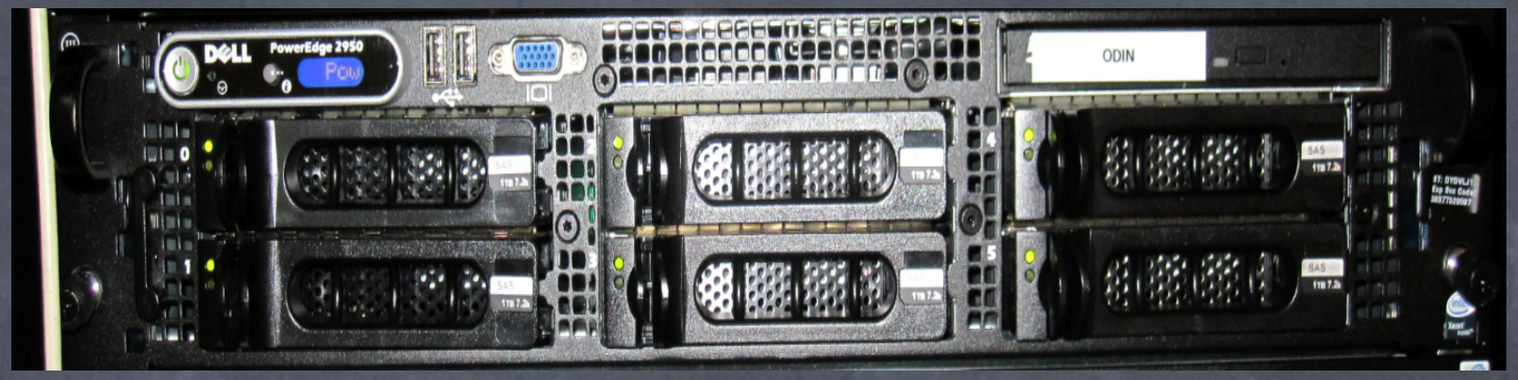
Infrastructure

- Archipelago
  - CAIDA's active measurement infrastructure
  - 43 monitors – growing 1 or 2 per month
  - 11 w/ Ipv6 connectivity
  - Team-probing experiment to collect v4 and v6 topology
  - Spoofer experiment

Infrastructure

- Passive Trace Capture
- Tier 1 OC192 backbone link packet header captures
- UCSD Network Telescope
  - 2 days of telescope dataset
    http://www.caida.org/data/passive/telescope-2days-2008_dataset.xml
  - 3 days of Conficker dataset
    http://www.caida.org/data/passive/telescope-3days-confi cker_dataset.xml

New infrastructure



- UCSD Telescope:
Intel Xeon 8 x 3GHz,
32GB RAM, 4.3TB

- Data Server:
Intel Xeon 8 x dual core
2.5GHz,16GB RAM,
8TB storage

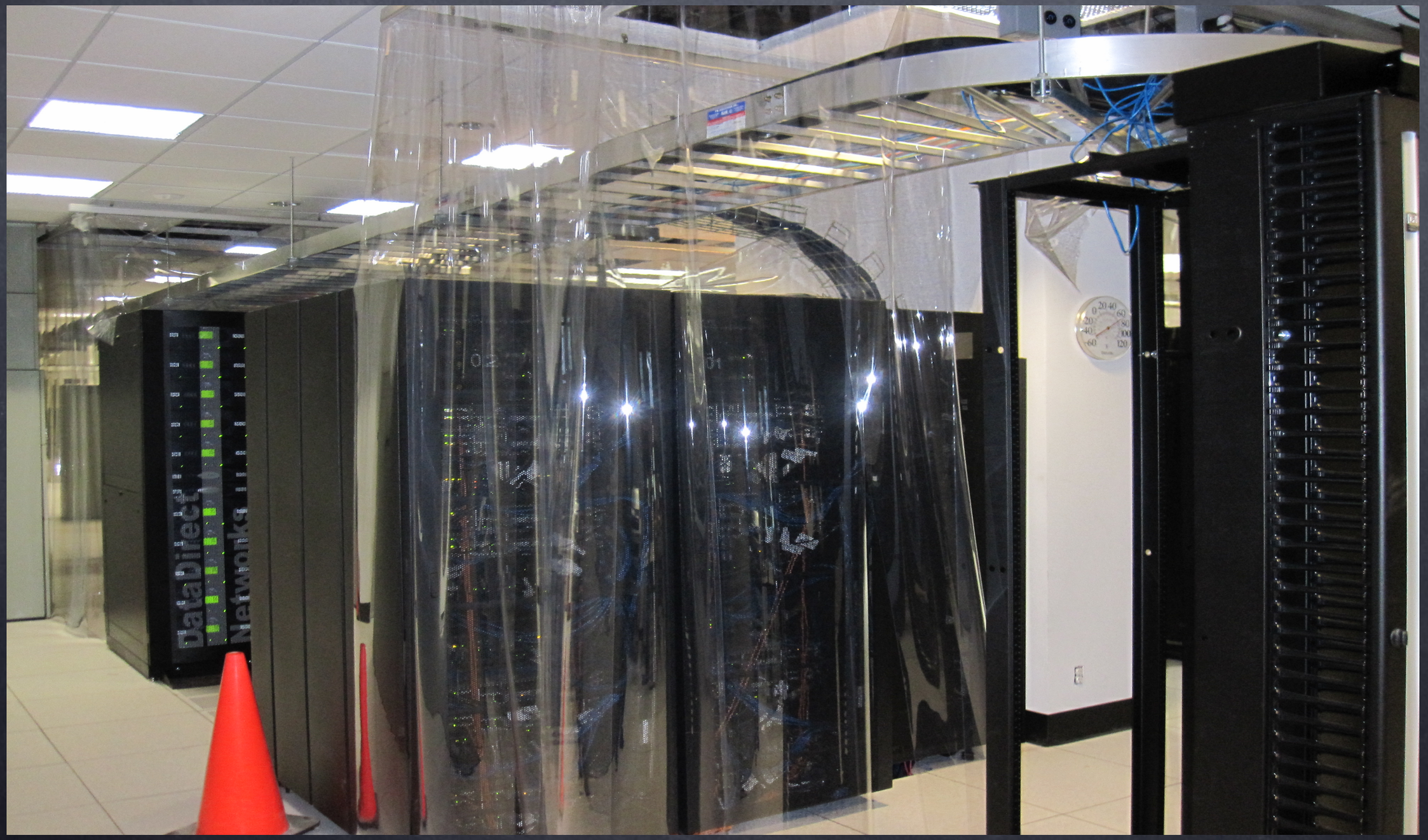
- New Ark Monitors
38 total active: 11 IPv6
(see topology project)

New infrastructure (cont)

- Doors and curtains

New infrastructure (cont)

- SDSC Machine room gets shower curtains



New infrastructure (cont)

- SDSC innovating in energy efficient machine room air handling systems where few standards exist
  - Old machine room focuses on hot aisles using curtains to force hot air to 16 air handlers. Allows us to increase air delivery set point from 55 degrees to 75 degrees. Curtains (from Subzero Coldrooms) have fuseable links that release in case of fire so fire systems will function.
  - New machine room focuses on cold aisles using homogenous racks to contain cold air to 500 sq/ft of the 5000 sq/ft space. Requires only 6 air handlers. (Knuerr)

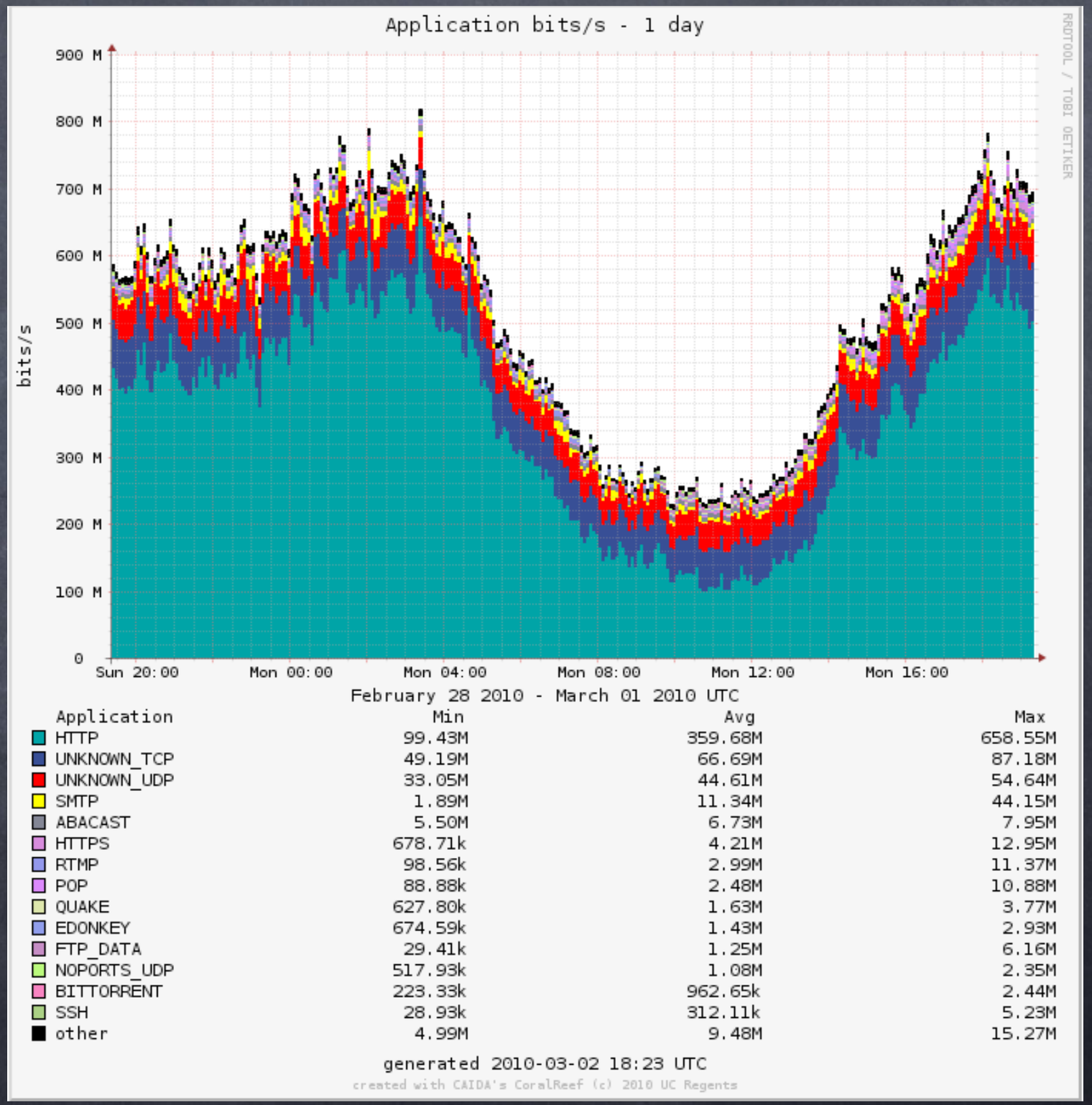
New machine room

Tools

- Coralreef
  - Report Generator
- Geocompare – survey of geolocation tools
- topostats – topology statistics
-

Tools (cont)

- Example: Report Generator - Chicago OC192 monitor



Data

- OC192 backbone: 8.5 TB (3.6 anonymized; 4.9 unanonymized) – curation to quarterlies will reduce
- UCSD telescope: 3.4 TB on disk (30 day window) 4.8 T on samqfs
- OC48 traces: 1.7TB (same old 2004 traces, in PREDICT)
- topology: 12.3 TB (skitter+ark uncompressed)
  - routed ipv4: 2.3TB (in PREDICT) since Sep 2007
  - routed ipv6: 275MB since Dec 2008

Total: ~30TB (as of 15 Feb 2010)

Data (cont)

- OC192 backbone: strip payload/L1/L2, transfer, anonymize, archive (aggregated links)
- OC48 traces: strip payload/L1/L2, anonymized w (prefix-preserve) cryptopan
- UCSD telescope: filter legitimate traffic at the router, 30 days on disk, curate custom data sets upon request
- topology: see cybersecurity project

(http://www.caida.org/home/legal)

How do researchers use the data?

- OC192 backbone: report generator up, traffic classification, performance modeling
  http://www.caida.org/data/publications/bydataset/index.xml#passive
- OC48 traces: traffic classification, modeling, monitoring, filtering, generation, locality
  http://www.caida.org/data/publications/bydataset/index.xml#OC48
- UCSD telescope: Conficker, worm research
  http://www.caida.org/data/publications/bydataset/index.xml#Backscatter
- topology: pkt traceback, marking. DOS defense. topo and routing modeling, discovery, metrics, improvements
  http://www.caida.org/data/publications/bydataset/index.xml#Topology

how do we use the data?

- OC192 backbone: traffic classification, real time monitor, traffic symmetry, address utilization, other myths
- OC48 traces: traffic classification, modeling, p2p, (also http://www.caida.org/data/realtime/passive/?monitor=sdnap)
- UCSD telescope: traffic classification, real-time monitor (/data/realtime/telescope), lots of (and not enough..) Conficker analysis
- Topology: annotated Internet mapping http://www.caida.org/research/topology/

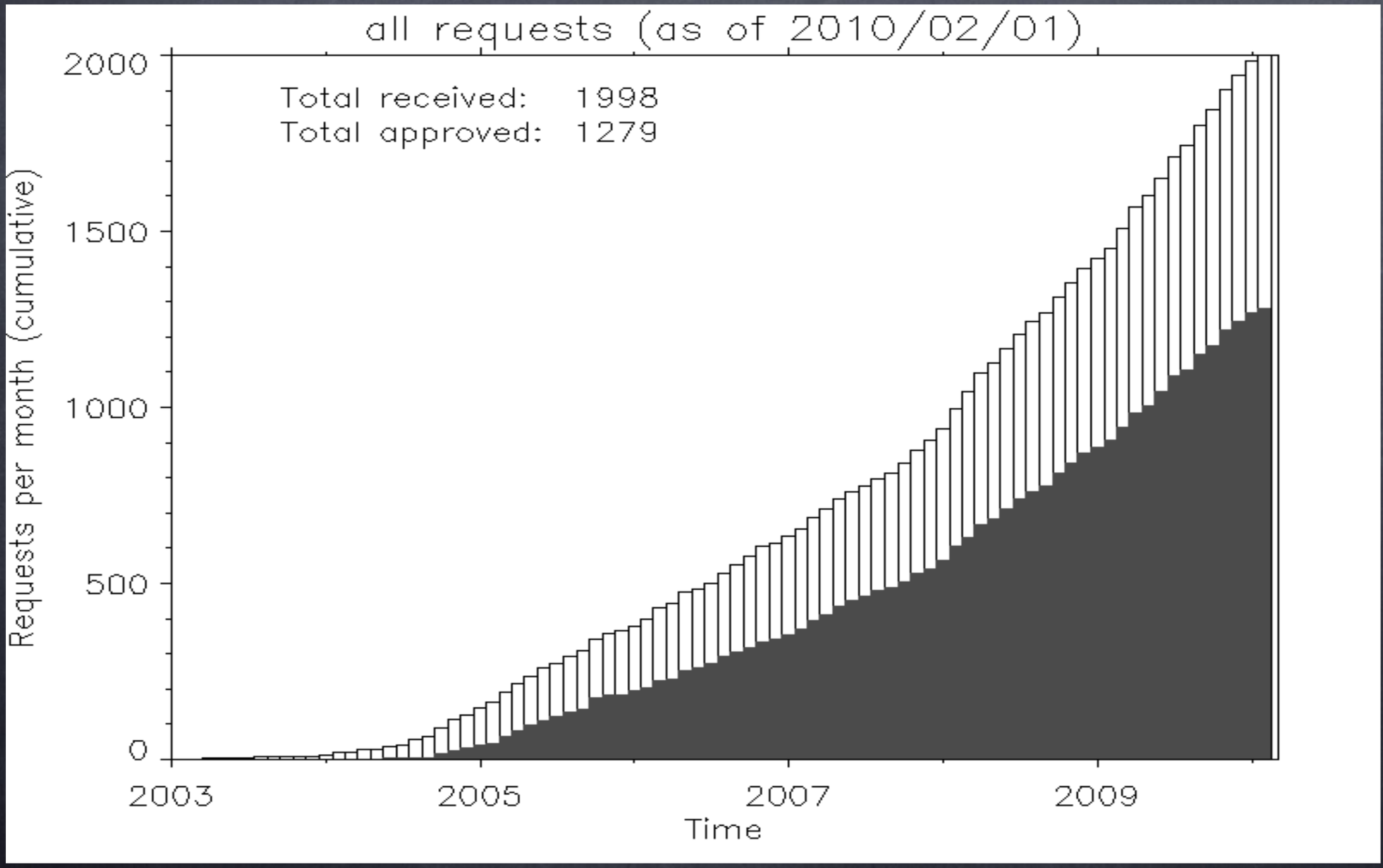
(www.caida.org/publications/papers/)

how many total requests for the data?

Dataset	Requests	Approved	Accessed	Since
Backscatter	451	241	207	Feb 2003
Passive	799	585	483	Feb 2004
Topology	614	372	290	Jul 2004
Witty	58	38	32	Mar 2008
Telescope	36	20	16	Jul 2009
DNS-RTT	40	23	18	Aug 2006
	1998	1279	1046	

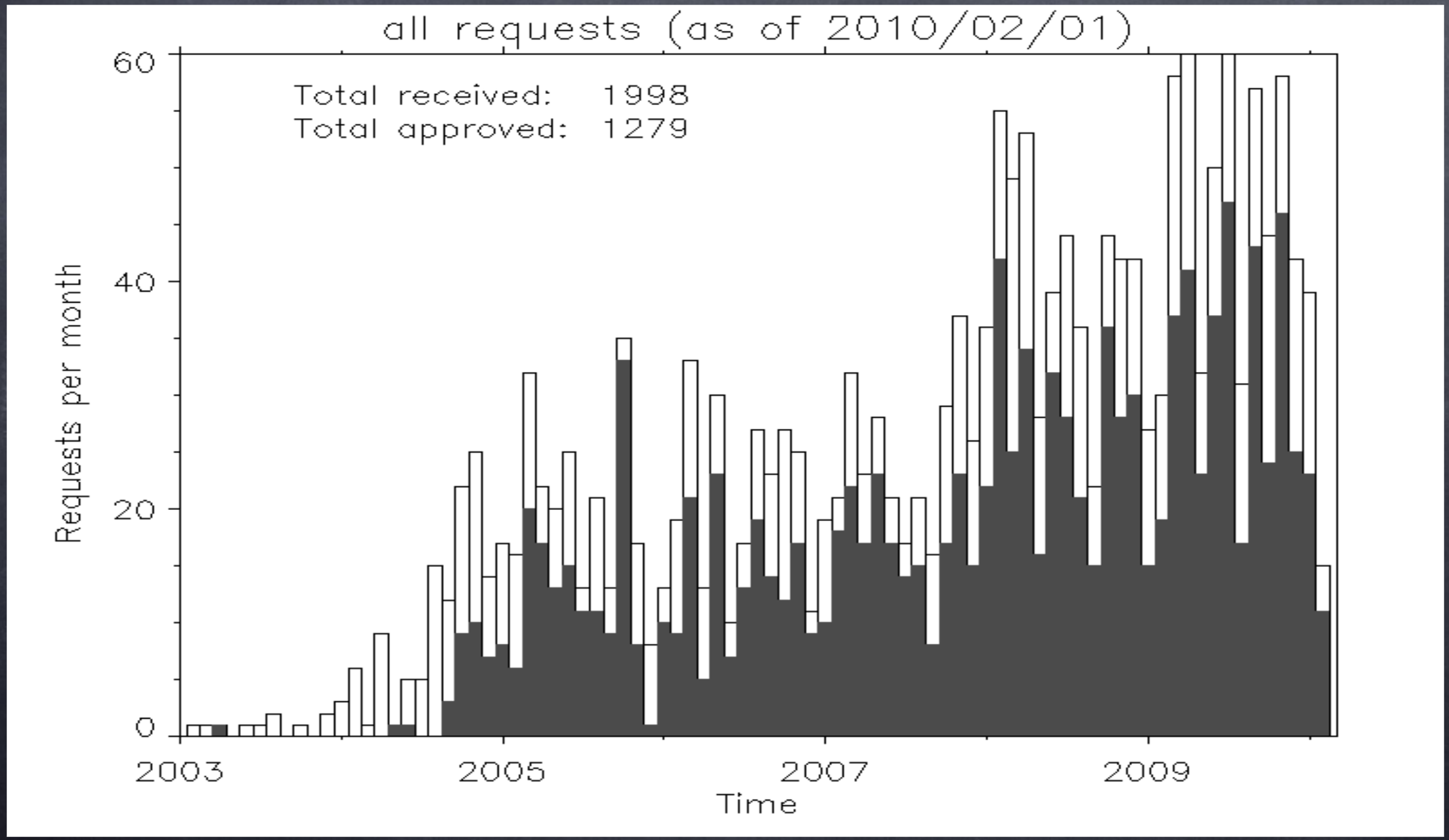
Data request stats

- All requests (cumulative)



Data request stats (cont)

- All requests (monthly)



Workshops

- Internet Statistics Metrics and Analysis (ISMA)
- Active Internet Measurement Systems (AIMS)
- Workshop on Internet Economics
- Joint workshops with OARC, WIDE, SFI

http://www.caida.org/workshops/

Potential Student Tasks

- Mathematical analysis of topology (sdp)
- Geolocation annotations of router-level graph using best available geo data
- Visualization of AS-level and router-level graphs
- Web site portal view of darkspace traffic

internships@caida.org