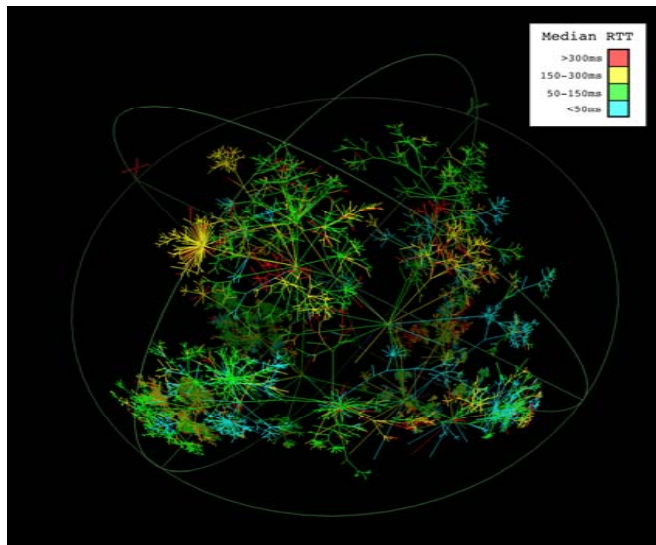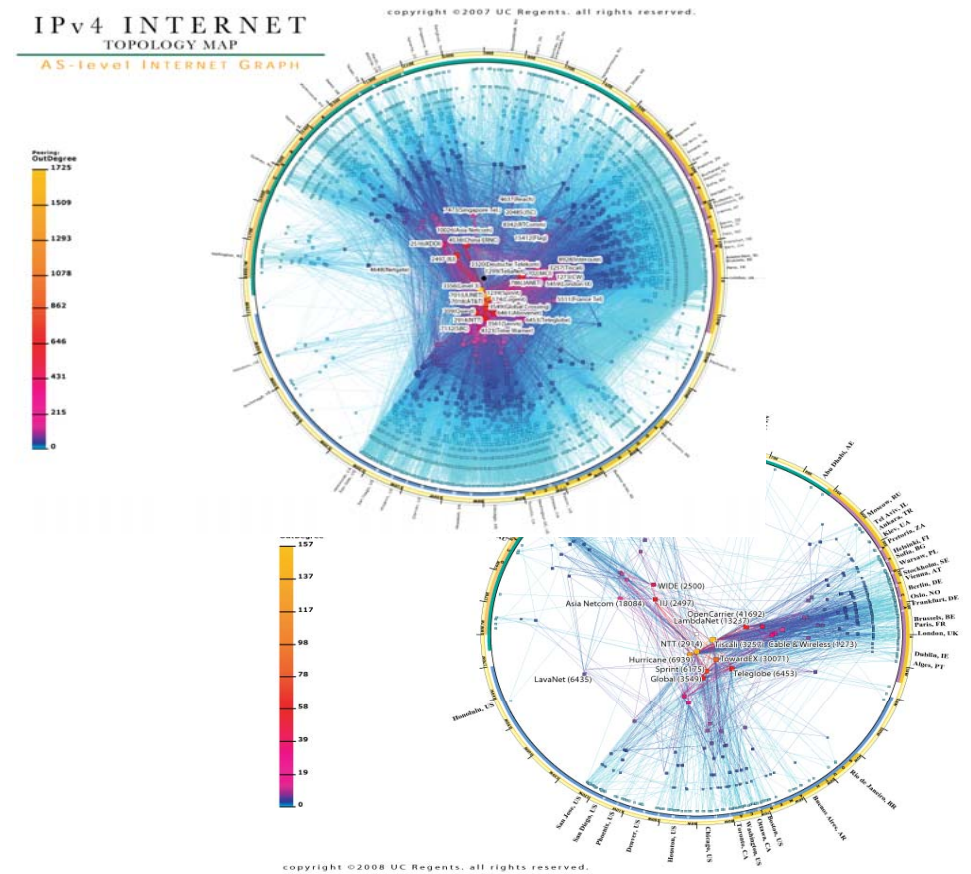# Internet Science: a survey of CAIDA activities

CAIDA = Cooperative Association for Internet Data Analysis

Marina Fomenkov, CAIDA

2nd CAIDA-WIDE-CASFI workshop

Seoul, April 4, 2009

Main projects

- Archipelago measurement infrastructure
- Topology measurements
- Routing theory
- Traffic analysis
- Policy and data sharing
- DNS
- Public outreach

Active measurement:*archipelago* (ark)

- replaces skitter
- 'operating system' for measurement
- launched 12 Sept 2007
- 32 active probers
- 7 are IPv6-capable
- future plans:
  - collaborators can run vetted measurements on security-hardened platform through simple API
  - general public can perform restricted measurements
  - support for meta-data mgt, analysis, and infoviz

3

Current experiments on Ark

- Ongoing IPv4 topology probing
- Beginning IPv6 topology probing
  - started in December 2008
- IP-to-router mapping
  - implementing improvements
- Spoofer (in collaboration with MIT)
  - general public is involved

Internet topology mapping with ark

- Supported by DHS Science & Technology directorate
- Integrate 6 strategic measurement & analysis capabilities for DHS 'situational awareness' needs:
  - new architecture for continuous topology measurements
  - IP alias resolution techniques
  - dual router- and AS-level graphs
  - AS taxonomy and relationships
  - geolocation of IP resources
  - graph visualization

Internet topology mapping with ark

- Milestones achieved:
  - Ongoing IPv4 topology measurements
  - Techniques for alias resolution
    - Iffinder, kapar
  - Router level graph
- Next steps
  - Dual router-AS level graph
  - Improved alias resolution
    - RadarGun
  - Automated data analysis and graph construction

Goal: regular publicly available graph updates

Internet topology mapping with ark

- To be continued…
  - Brad's presentation

- Papers in preparation:
  - Alias Resolution techniques and results
  - Ark probing tools and methods
  - spoofer

New approach to routing

- The ultimate problem with routing scalability is the *updates*

  can we route without updates?

- Structure of observed complex networks (strong clustering, specific power laws) maximizes their navigability

- Mathematical approach: *hidden metric space* underlying the observed network toplogy

Hidden Metric Spaces

QuickTime™ and a
decompressor
are needed to see this picture.

- Triangle inequality explains strong clustering
- Guides greedy routing process
- Negative curvature - hyperbolic

Hidden Metric Spaces

- ## PI Dima Krioukov
  - Work in progress
- ## Broad impact for other disciplines
  - Social, biological, neural networks
- ## Publications:
  - Self-similarity of complex networks and hidden metric spaces, *Phys.Rev.Let.*, Apr 2008
  - Navigability of complex networks, *Nature Physics*, Jan 2009
  - Navigating ultrasmall worlds in ultrashort time, *Phys.Rev.Let.*, Feb 2009

Passive measurements

- Historically, one of the CAIDA main interests
- Severely hindered by the lack of coherent privacy policies
- Traffic monitor at Equinix data center in Chicago connected to an OC192 backbone link
  - Monthly traces, 1 hour long
  - Will change to quarterly traces
- Traffic monitor at Equinix data center in San Jose
  - Not fully operational yet

Coral Reef: software for traffic analysis

- collects and analyze data from passive Internet traffic monitors, in real time or from trace files.

- programming APIs for C, Perl; applications for capture, analysis, and web report generation.

- CAIDA developers maintain with help from Internet measurement community.

  http://www.caida.org/tools/measurement/coralreef/

UCSD Network telescope

- Planned to turn it off
- Conficker happened!
  - Also known as Downadup, Conflicker, Kido.
  - Monitoring Conficker's TCP scanning behavior
    - searching for victim machines to exploit
  - Observed Conficker.A and .B versions
  - www.caida.org/research/security/ms08-067/conficker.xml
- April 1st - ?

Data sharing and Policy

- DatCat - Internet Measurement Data Catalog
  - unfunded, volunteer efforts
- PREDICT - funded by DHS
  - Data Provider, Data Host, Advisory role
- kc's Blog
  - http://blog.caida.org/best_available_data/
- IRB - Institutional Review Board
  - Required for human objects research

DNS research

- Running out of funding in 2009
- Main focus:
  - DITL measurements
  - Data analysis
  - Simulations
- DITL 2009: March 31- April 1
  - A, C, E, F, H, K, L, M root servers are participating
- Measurements for new gTLDs impact?

DNS Measurements

Mean query rate at the root servers (2006, 2007, 2008)
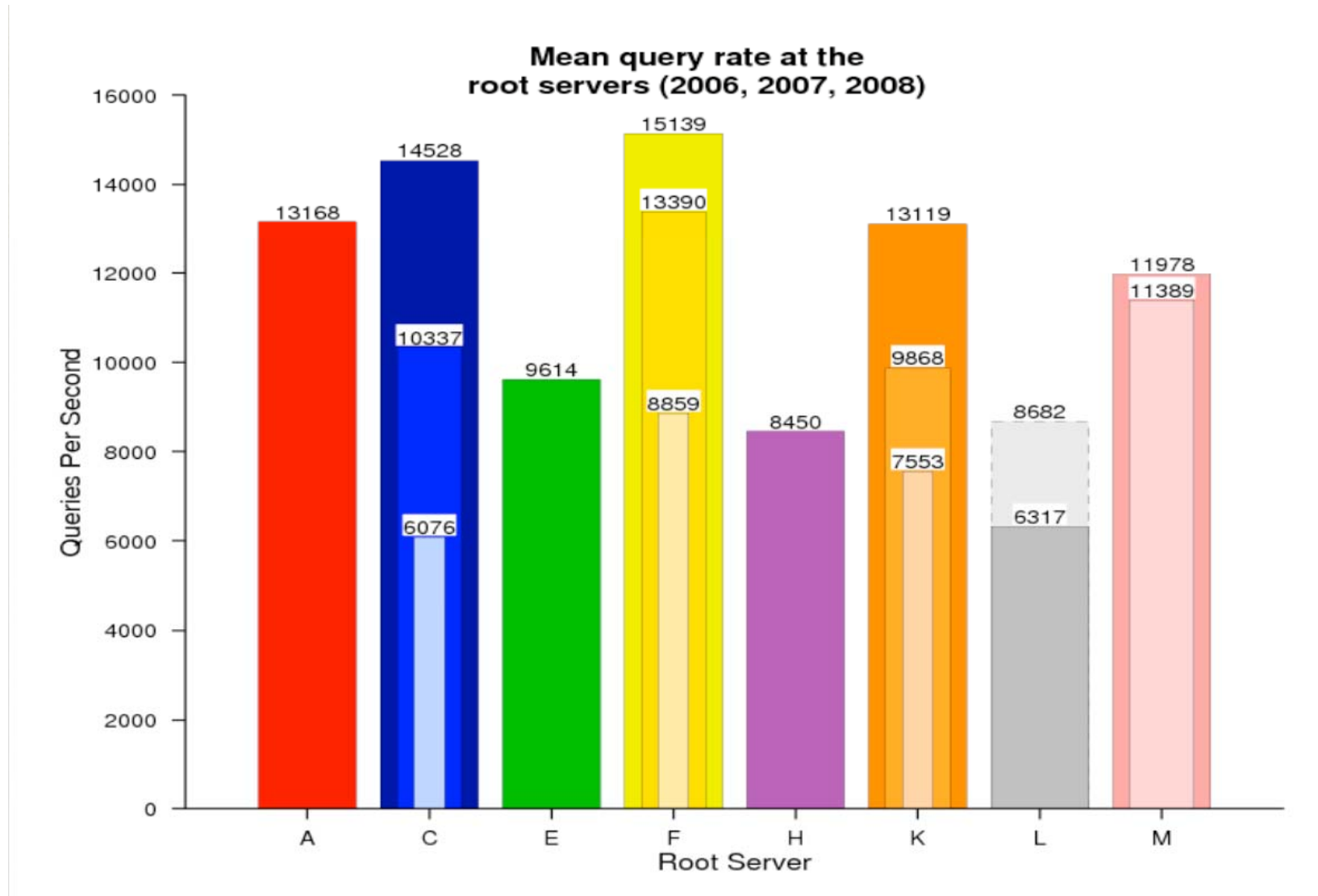
Public outreach

- Education
  - Students, interns, postdocs
  - Teaching at UCSD
- Workshops
  - Active Internet Measurements (AIMS) in support of Internet science and policy, Feb 09
- Publications and presentations
- Blog

CAIDA future

- Bright, yet uncertain
- Exciting, cutting edge research projects
- Looking for funding
  - NSF
  - DHS
  - Members and gifts
- Looking for postdocs