CAIDA Activities

TERENA – May 22, 2007

Colleen Shannon

cshannon@caida.org

What is CAIDA?

- Cooperative Association for Internet Data Analysis
- http://www.caida.org/

- Main Goal: measuring and understanding the global Internet
- Develop measurement and analysis tools
- Collect and provide Internet data: topology, header traces, routing, network security, DNS
- Visualization of the network

Outline

- Macroscopic Topology Measurement
- Routing
- DNS
- Security
- COMMONS
- Data Collection, Curation, and Distribution
  - DatCat: the Internet Measurement Data Catalog
- Tools

Macroscopic Topology Measurement (Brad Huffaker, Young Hyun)

- Skitter project continues: daily traceroute-like measurements from 22 monitors to locations
- New active measurement infrastructure: Archipelago (Ark)
  - Architecture supports:
    - Coordinated measurements (e.g. team probing)
    - Shared use of a common measurement infrastructure
    - Security measures to ensure responsible use and data integrity

Routing
(Dima Krioukov)

- Realistic topology generation
  - dK series graphs can generate topologies that satisfy a series of graph properties
- AS Adjacencies
  - Traceroute-based matrix of Internet AS-level graph
- AS Relationships, Classification, and Taxonomy
  - AS adjacencies annotated with information such as inferred customers and providers and IP address space

Domain Name System (DNS) (Duane Wessels, Marina Fomenkov)

- DNS Surveys
  - Open resolvers (recursive name resolution to folks outside their administrative domain)
  - Cache poisoning (incorrect referrals for important domains)
  - Nameserver software prevalence
- RTT measurements to DNS root and gTLD servers

Current Security Research (David Moore, Colleen Shannon)

- Nyxem/Blackworm/KamaSutra/MyWife
  - http://www.caida.org/analysis/security/blackworm/
- Spamscatter
- Botnet Economics
- Worm Risk Analysis
- Anomaly Detection

COMMONS
(k claffy)

- Cooperative Measurement and Modeling of Open Networked Systems
- Problems:
  - Infrastructure financial crisis
  - Data acquisition crisis
  - Struggle for survival for emerging community/municipal wireless network
- Solution: Cooperative national backbone connecting community and municipal networks
  - Low-cost access for community wireless networks via shared network resources
  - Implicit support (and consent) for measurement activities

Day in the Life of the Internet

- At-least annual measurement with as many networks participating as possible
- Most recent: January 9-10, 2007
  - 7 DNS participants (C root, F root, K root, M root, AS112, B ORSN, M ORSN)
  - 5 network participants (WIDE, KAIST, POSTTECH, AMPATH, CAIDA)
- To join future DITL data collections, email ditl-info@caida.org

CAIDA Datasets

- Freely available datasets
- Academic / Non-profit access datasets
- For-profit use: sponsor dataset creation
  - Join CAIDA: http://www.caida.org/home/legal/sponsorinfo.xml
  - US organizations: use PREDICT
    - http://www.predict.org/

Freely Available Data

- The following datasets are available to anyone who wishes to use them:
  - AS Adjacencies
  - Router Adjacencies
  - Code-Red Worm
  - Witty Worm
  - AS Relationships
  - AS Rank
  - AS Taxonomy

Data available for non-profit use

- The following datasets are available to academic, government, and non-profit researchers:
  - Raw macroscopic topology traces (skitter)
  - OC48 peering point data
  - Denial-of-service attack backscatter (TOCS, 2004-2005, 2006)
  - Witty Worm
  - DNS root/gTLD RTT data

Internet Measurement Data Catalog

http://imdc.datcat.org

DatCat Goals (1)

- to facilitate searching for and sharing of data among researchers
  - Index as much as possible, including datasets not publicly available
  - DatCat doesn't store any network data itself

DatCat Goals (2)

- to enhance documentation of datasets via a public annotation system
  - Easy place for anyone (not just the dataset creator) to provide additional information
  - Persistent reference that stays with the dataset (not a footnote in a paper)

DatCat Goals (3)

- to advance network science by promoting reproducible research
  - Test new technologies on consistent datasets to compare apples with apples

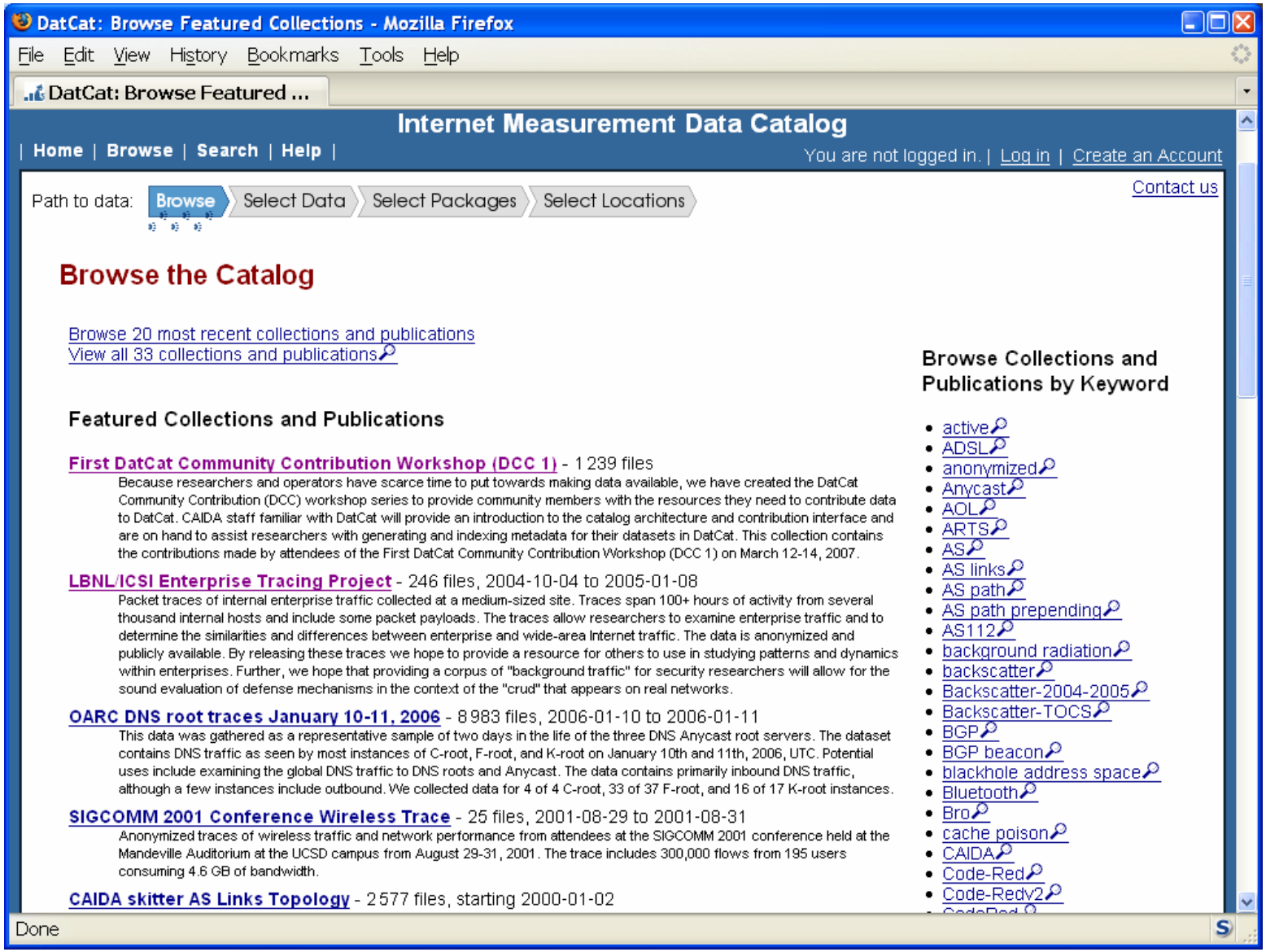
DatCat lets you…

- Find data for research/engineering
- Annotate datasets to note features, background information, or bugs
- Cite data
- Contribute data (coming soon!)

DatCat Status

- DatCat available for public viewing since June 12, 2006
- Contribution interface open to beta-testers
- 76,708 data items
- 6 TB of data
- 33 Collections and Publications
  - 15 non-CAIDA Data Collections (26 total)
  - 6 non-CAIDA Publications (7 total)

DatCat Example



Collaboration

- Current:
  - CRAWDAD: Community Resource for Archiving Wireless Data at Dartmouth
  - MOME/MOMENT
  - UCSD-CSE, ICSI
- Future:
  - Abilene Observatory
  - RouteViews

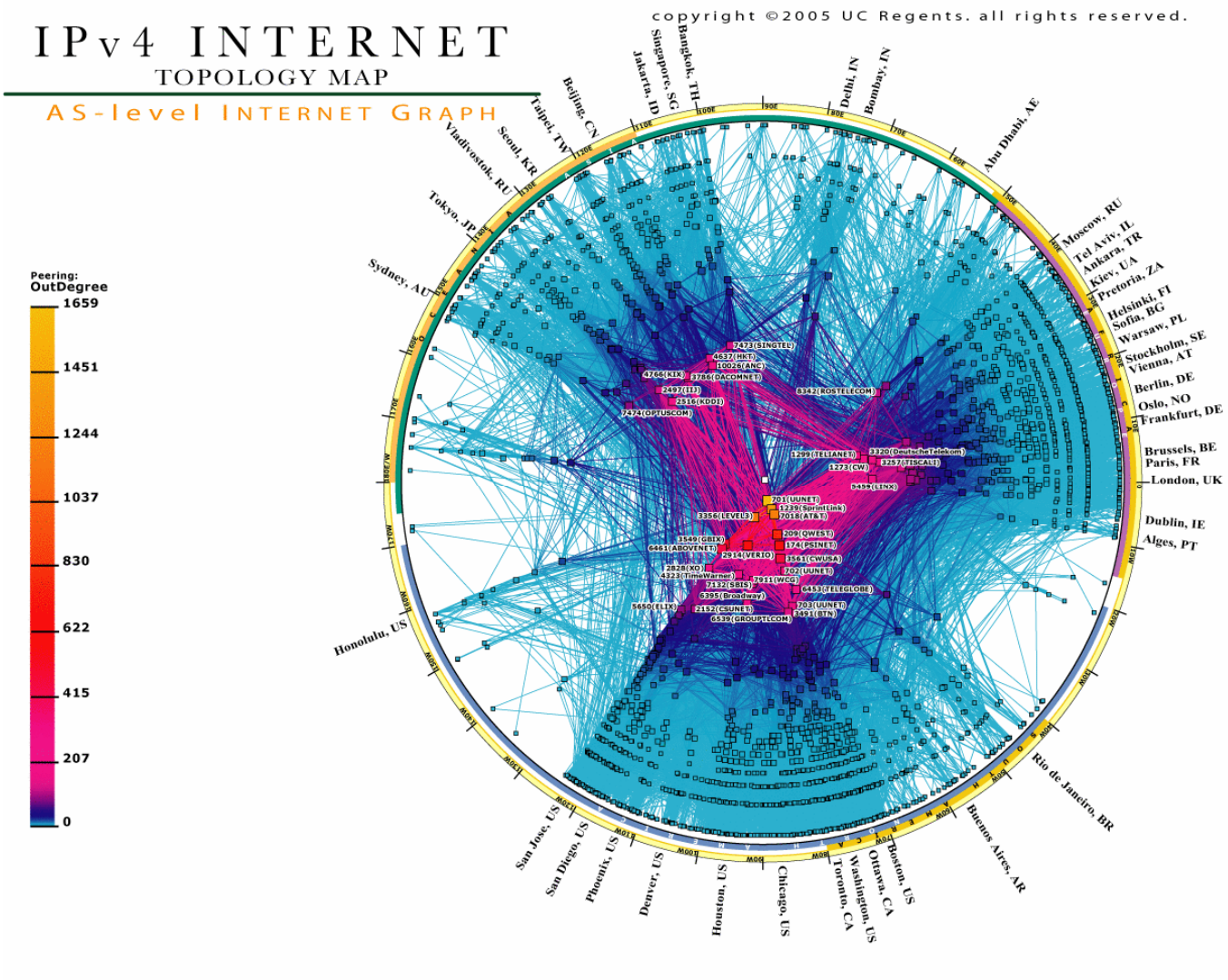
For more information

- DatCat: http://imdc.datcat.org/
- General questions and comments
  - info@datcat.org
- Announcements
  - user-announce@datcat.org
- Contribution beta-test
  - contribute@datcat.org

CAIDA Tools

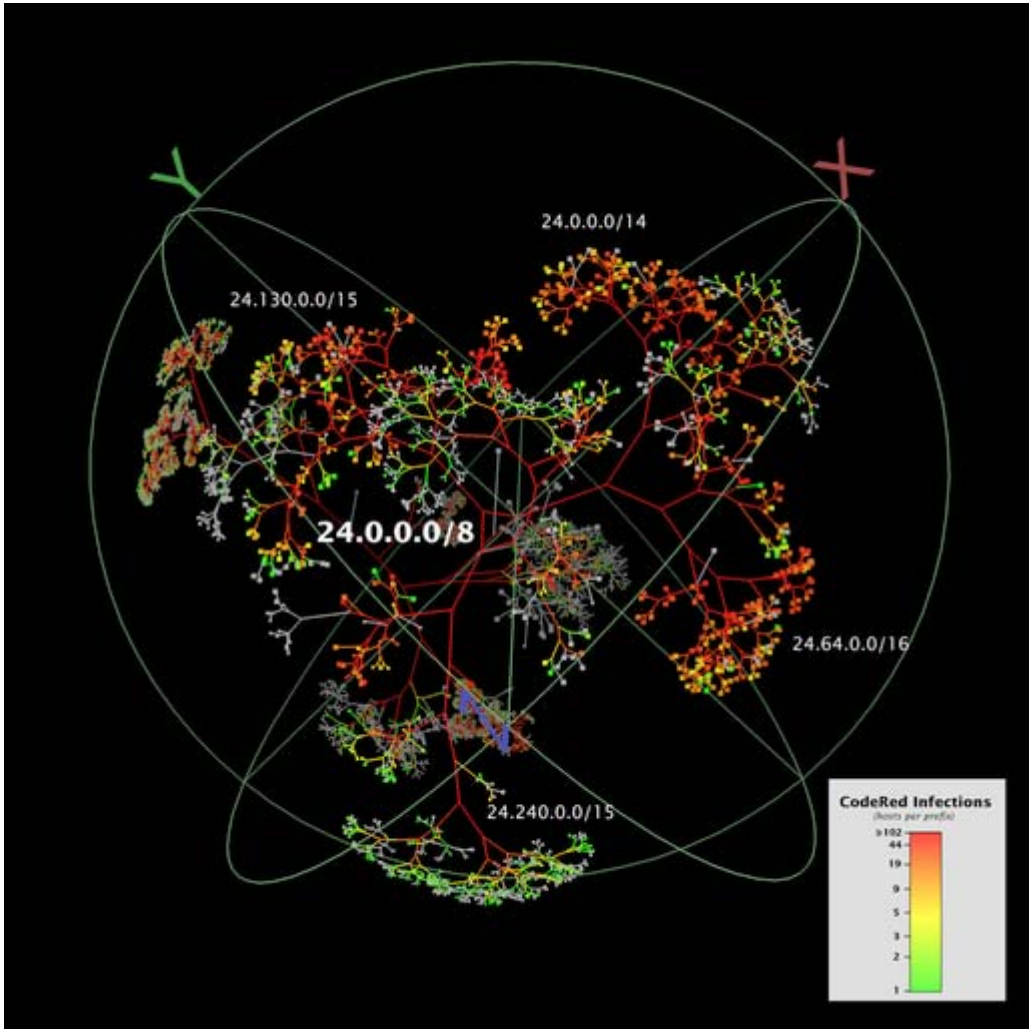
- Measurement and analysis
  - CoralReef
  - Scamper
  - NeTraMet
  - DSC
- Visualization
  - Walrus
  - Cuttlefish
  - Otter

Otter Example: AS Connectivity Map



caida

Walrus Example: Code-Red Worm



Cuttlefish Example: Blackworm Virus

- (live demo)

For more information…

- CAIDA Research:
  - http://www.caida.org/research/
- CAIDA Data:
  - http://www.caida.org/data/
- DatCat:
  - http://imdc.datcat.org
- CAIDA Tools
  - http://www.caida.org/tools/

Contact Information

- Questions about this talk:
  - cshannon at caida.org
- Questions about CAIDA in general
  - Info at caida.org
- Questions about CAIDA data
  - Data-info at caida.org
- Questions about the Day in the Life of the Internet (DITL) project
  - Ditl-info at caida.org
- Questions about DatCat
  - Info at datcat.org
  - Contribute at datcat.org