

# CAIDA Activities

TERENA – May 22, 2007

Colleen Shannon

[cshannon@caida.org](mailto: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](mailto: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>



---

Cooperative Association for Internet Data Analysis

# 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

DatCat: Browse Featured Collections - Mozilla Firefox

File Edit View History Bookmarks Tools Help

DatCat: Browse Featured ...

Internet Measurement Data Catalog

Home | Browse | Search | Help | You are not logged in. | Log in | Create an Account

Contact us

Path to data: Browse > Select Data > Select Packages > Select Locations

## Browse the Catalog

[Browse 20 most recent collections and publications](#)  
[View all 33 collections and publications](#)

### Featured Collections and Publications

**First DatCat Community Contribution Workshop (DCC 1)** - 1 239 files  
Because researchers and operators have scarce time to put towards making data available, we have created the DatCat Community Contribution (DCC) workshop series to provide community members with the resources they need to contribute data to DatCat. CAIDA staff familiar with DatCat will provide an introduction to the catalog architecture and contribution interface and are on hand to assist researchers with generating and indexing metadata for their datasets in DatCat. This collection contains the contributions made by attendees of the First DatCat Community Contribution Workshop (DCC 1) on March 12-14, 2007.

**LBNL/ICSI Enterprise Tracing Project** - 246 files, 2004-10-04 to 2005-01-08  
Packet traces of internal enterprise traffic collected at a medium-sized site. Traces span 100+ hours of activity from several thousand internal hosts and include some packet payloads. The traces allow researchers to examine enterprise traffic and to determine the similarities and differences between enterprise and wide-area Internet traffic. The data is anonymized and publicly available. By releasing these traces we hope to provide a resource for others to use in studying patterns and dynamics within enterprises. Further, we hope that providing a corpus of "background traffic" for security researchers will allow for the sound evaluation of defense mechanisms in the context of the "crud" that appears on real networks.

**OARC DNS root traces January 10-11, 2006** - 8 983 files, 2006-01-10 to 2006-01-11  
This data was gathered as a representative sample of two days in the life of the three DNS Anycast root servers. The dataset contains DNS traffic as seen by most instances of C-root, F-root, and K-root on January 10th and 11th, 2006, UTC. Potential uses include examining the global DNS traffic to DNS roots and Anycast. The data contains primarily inbound DNS traffic, although a few instances include outbound. We collected data for 4 of 4 C-root, 33 of 37 F-root, and 16 of 17 K-root instances.

**SIGCOMM 2001 Conference Wireless Trace** - 25 files, 2001-08-29 to 2001-08-31  
Anonymized traces of wireless traffic and network performance from attendees at the SIGCOMM 2001 conference held at the Mandeville Auditorium at the UCSD campus from August 29-31, 2001. The trace includes 300,000 flows from 195 users consuming 4.6 GB of bandwidth.

**CAIDA skitter AS Links Topology** - 2 577 files, starting 2000-01-02

### Browse Collections and Publications by Keyword

- active
- ADSL
- anonymized
- Anycast
- AOL
- ARTS
- AS
- AS links
- AS path
- AS path prepending
- AS112
- background radiation
- backscatter
- Backscatter-2004-2005
- Backscatter-TOCS
- BGP
- BGP beacon
- blackhole address space
- Bluetooth
- Bro
- cache poison
- CAIDA
- Code-Red
- Code-Redv2
- Code-Redv3

Done



# 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](mailto:info@datcat.org)
- Announcements
  - [user-announce@datcat.org](mailto:user-announce@datcat.org)
- Contribution beta-test
  - [contribute@datcat.org](mailto:contribute@datcat.org)



# CAIDA Tools

---

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

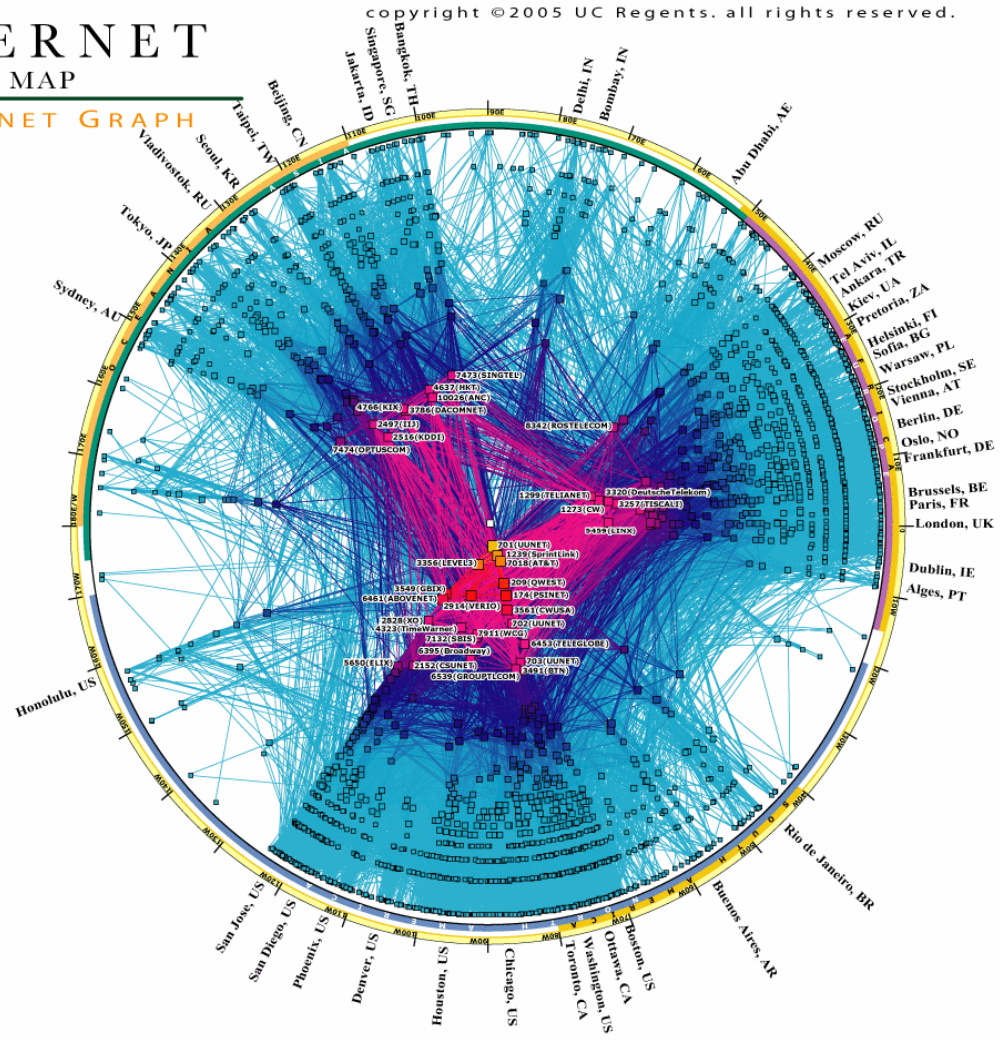
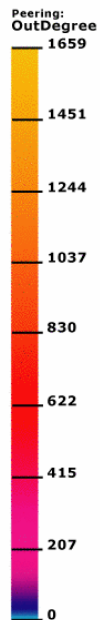


# Otter Example: AS Connectivity Map

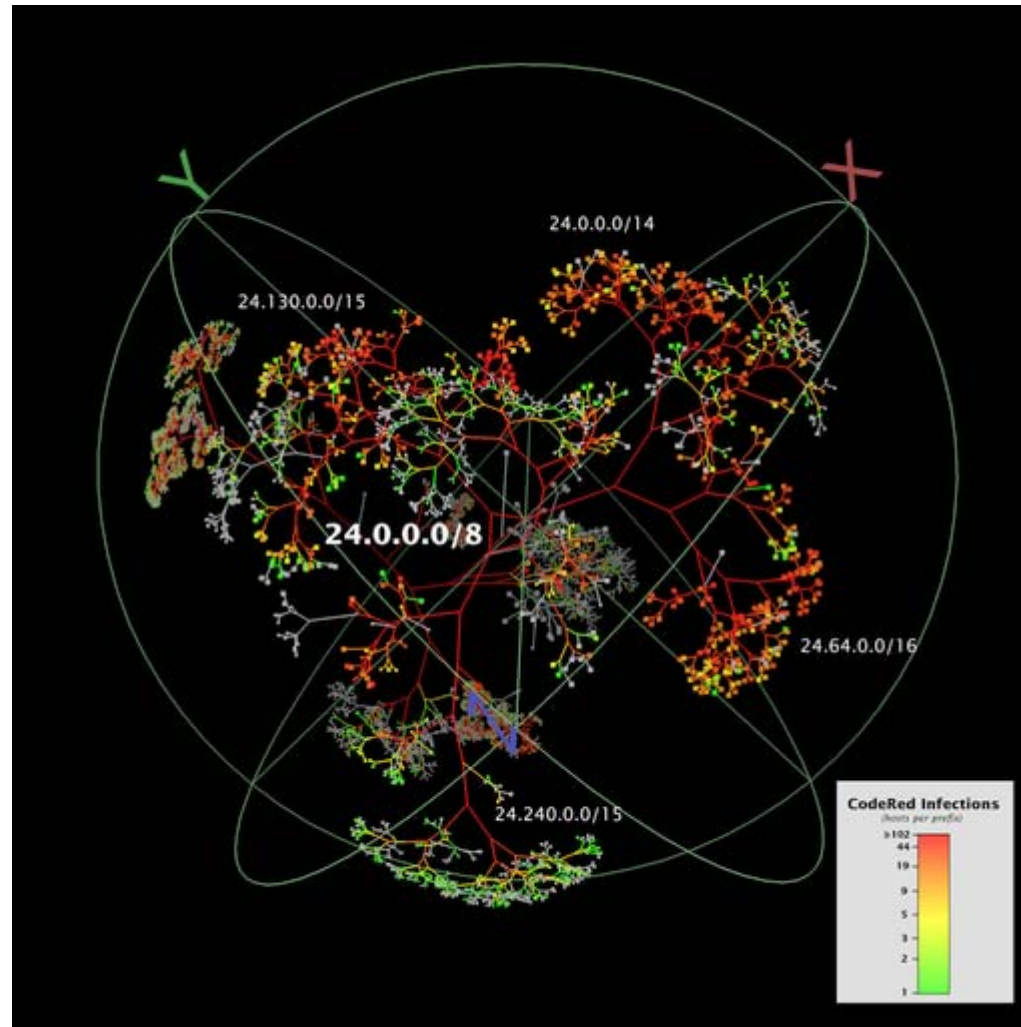
## IP<sub>v</sub>4 INTERNET TOPOLOGY MAP

### AS-level INTERNET GRAPH

copyright ©2005 UC Regents. all rights reserved.



# 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](mailto:cshannon@caida.org)
- Questions about CAIDA in general
  - [Info at caida.org](http://caida.org)
- Questions about CAIDA data
  - [Data-info at caida.org](http://caida.org/data-info)
- Questions about the Day in the Life of the Internet (DITL) project
  - [Ditl-info at caida.org](http://caida.org/ditl-info)
- Questions about DatCat
  - [Info at datcat.org](http://datcat.org)
  - [Contribute at datcat.org](http://datcat.org/contribute)

