

DITL 2007 Collection Summary

Josh Polterock <josh@caida.org>

CAIDA/WIDE Workshop

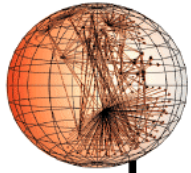
January 19, 2008

Honolulu, HI

USA



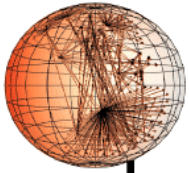
<http://www.caida.org/>



caida Overview

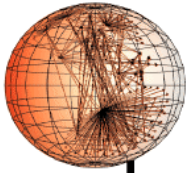
DITL January 9-10, 2007 Summary

- Motivation
- The Participants & The Data
- Data Access
- The MetaData
- Conclusion
- Related Links



caida Motivation

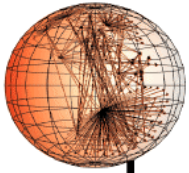
- To capture a day in the life of the Internet with as much scientifically grounded methodology as possible, and with resulting data as widely accessible as possible.
- Goals:
 - To build community and infrastructure for documentation.
 - To coordinate regular data collection events.



caida

The Participants & The Data

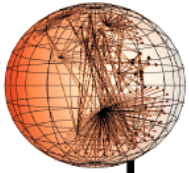
- OARC DNS Root Nameservers
 - c.root (Cogent), e.root (NASA), f.root (ISC), k.root (RIPE NCC), m.root (WIDE)
 - Inbound/outbound DNS packet traces with payload
 - no outbound from f.root; no data from e.root
- NaMeX Internet Exchange
 - AS112.nameex.it authoritative DNS servers for the reverse zones RFC3330 address space
- Open Root Server Network (ORSN)
 - b.orsn (FunkFeuer Free Net), m.orsn (Home of the Brave GmbH)
 - Inbound/outbound DNS packet traces with payload



caida

The Participants & The Data (cont.)

- Widely Integrated Distributed Environment (WIDE)
 - CAMPUS 10 Gigabit Ethernet Trace
 - TRANSIT 1 Gigabit Ethernet Trace (available anonymized)
- Korea Advanced Institute of Science and Technology (KAIST)
 - KOREN 1 Gigabit Ethernet Trace
 - Hanaro 1 Gigabit Ethernet Trace
- Pohang University of Science and Technology (POSTECH)
 - 1 Gigabit Ethernet Trace
- Chungnam National University (CNU)
 - 1 Gigabit Ethernet Trace
- AMPATH (AMericaSPATH): International Exchange Point in Miami
 - Anonymized OC12 ATM Trace



caida

The Participants & The Data (cont.)

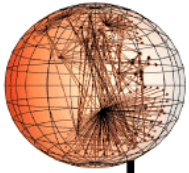
- CAIDA
 - Skitter traceroute topology data
 - Denial of Service Backscatter data
 - IRCache Squid web cache logs
- Other datasets
 - Internet2 Netflow
 - U of Oregon RouteViews



Coverage Over Collection Period

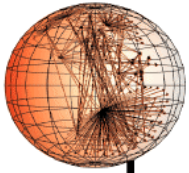
- Example of coverage of DNS root nameserver datasets





caida Data Access

- OARC DNS Data - c.root, f.root, k.root, m.root, ORSN, and AS112
 - Restricted: Access via OARC Access Agreement requires that the data and analysis must remain on OARC servers.
- WIDE
 - Unrestricted: Anonymized versions w/o payload publicly available
- POSTECH-KT
 - Restricted: contact via email for access
- Korea Advanced Institute of Science and Technology-KOREN
 - Restricted: access requests accepted for collaborative agreements to share analysis, implementation code, and results.

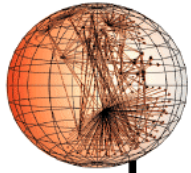


caida

Data Access (continued)

DITL January 9-10, 2007 Summary

- Korea Advanced Institute of Science and Technology-Hanaro
 - Restricted: researchers may submit analysis code for KAIST staff to run on data
- Chungnam National University
 - Restricted: researchers may submit analysis code for CNU staff to run on data
- CAIDA
 - Restricted: available to academic, government and non-profit researchers and members upon request

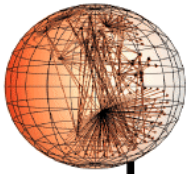


caida

Increased Access to Data

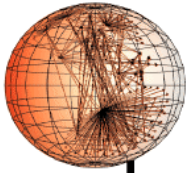
- Though most have access restrictions, these datasets are indexed in the Internet Measurement Data Catalog (IMDC) nicknamed “DatCat.”
- DatCat provides metadata and annotations and information about raw data availability.
- Researchers can learn about datasets before they request direct access.





caida The Metadata

- Documents methodology including specific options and local machine configurations (e.g. tcpdump)
 - basic attributes: filenames, sizes, long and short descriptions, keywords, md5 hash, etc.
 - further details: anomalies or peculiarities in the data, clock skew, etc...

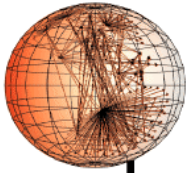


caida

Future Areas of Work

DITL January 9-10, 2007 Summary

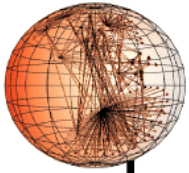
- Complete indexing the remaining datasets
- This year, we hope:
 - to finish indexing more promptly,
 - recruit more participants, and
 - cover more data types.



caida

Conclusions

- The 2007 collection event enjoyed:
 - twice the number of participants,
 - a larger number of data types,
 - to our knowledge, the broadest coordinated Internet collection event ever.



caida Links

- A Day In The Life of the Internet
<http://www.caida.org/projects/ditl/>
- A Day In The Life of the Internet:
A Summary of the January 9-10, 2007 Collection Event
<http://www.caida.org/projects/ditl/summary-2007-01/>
- Internet Measurement Data Catalog (DatCat)
<http://imdc.datcat.org/>
- Day in the Life of the Internet, January 9-10, 2007 (DITL-2007-01-09)
<http://imdc.datcat.org/collection/1-031B-Q>