

Two days in the life of
three DNS root servers

+



overview and demonstration

cooperative association for internet data analysis

Bradley Huffaker <bradley@caida.org>

16 November 2006



DNS anycast analysis

**Using tcpdump from three roots
we examined the geographic and topological
clustering of DNS clients.**

**Data collected by OARC ISC with
COGENT and RIPE NCC.**



Outline

DNS anycast analysis

- data source
- analysis
 - ◆ diurnal patterns
 - ◆ num. of requests/addresses per instance
 - ◆ geographic relationships / distances
 - ◆ coverage topological / geographic
- visualization
 - ◆ Influence Map

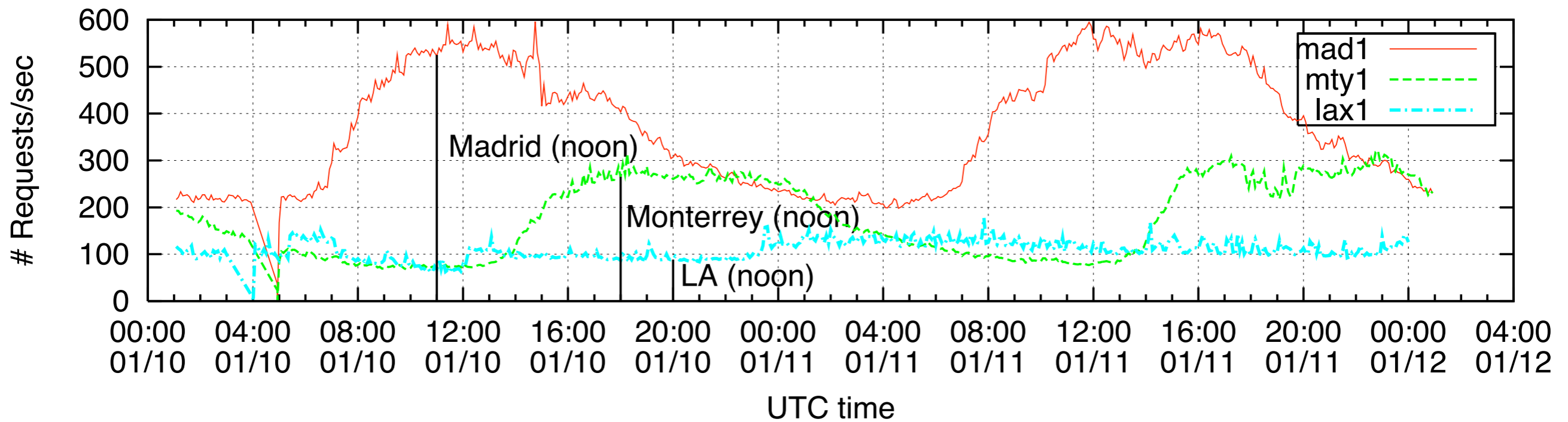


Data Source

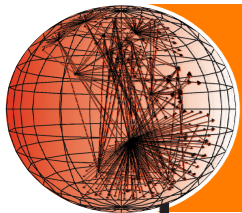
- dates
 - January 10th-11th 2006 (47 hours)
- DNS sources
 - c-root (Cogent) 4 instances
 - f-root (ISC) 32 instances
 - k-root (RIPE) 11 instances
- geographic
 - Netacuity database used for geographic mapping
- topological
 - Route Views used for ASs and prefixes (January 10th)



Diurnal Patterns

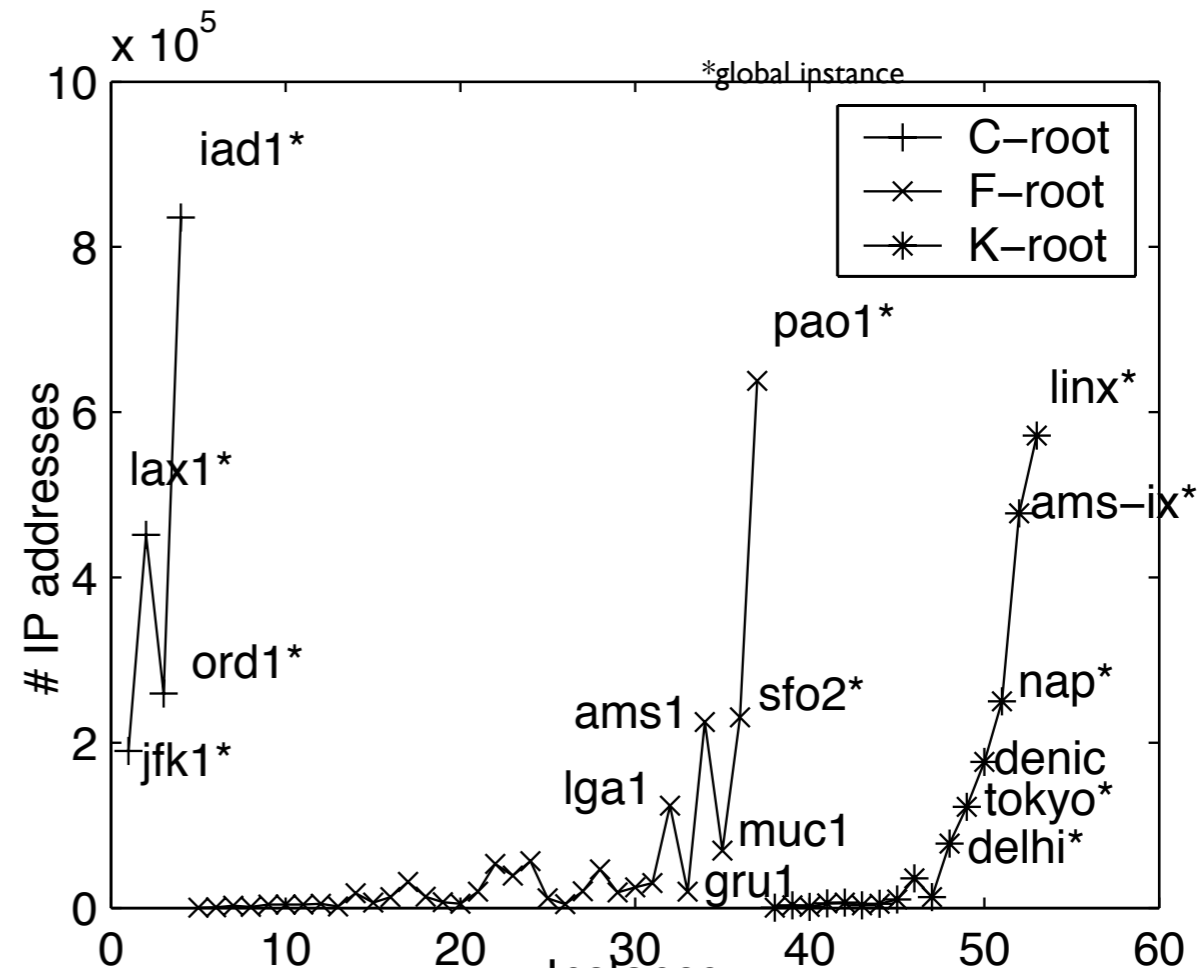
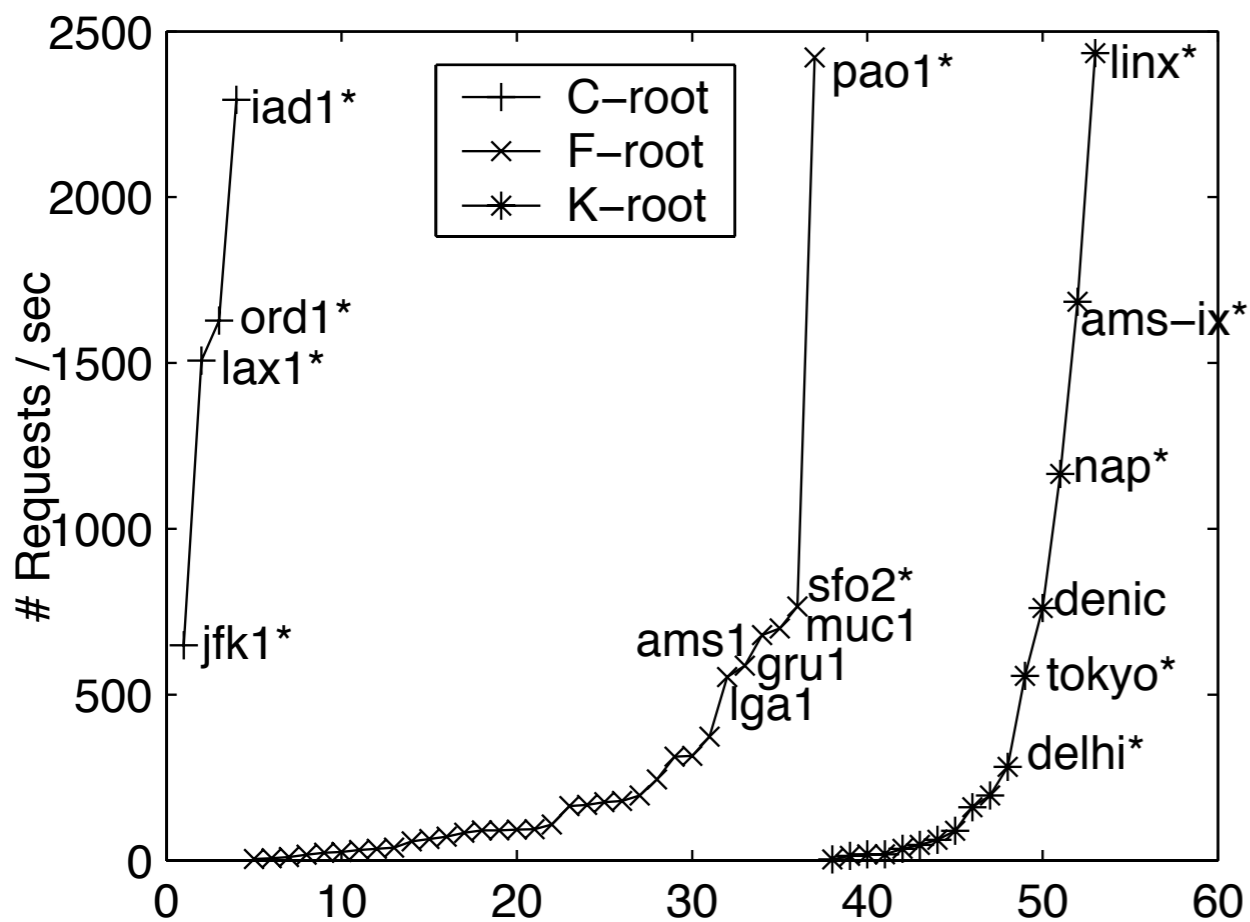


The local instances mad1 and mty1 show clear diurnal patterns

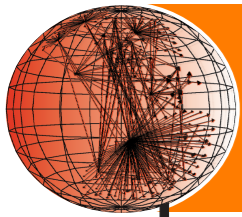


caida

Num. of Requests/Addresses

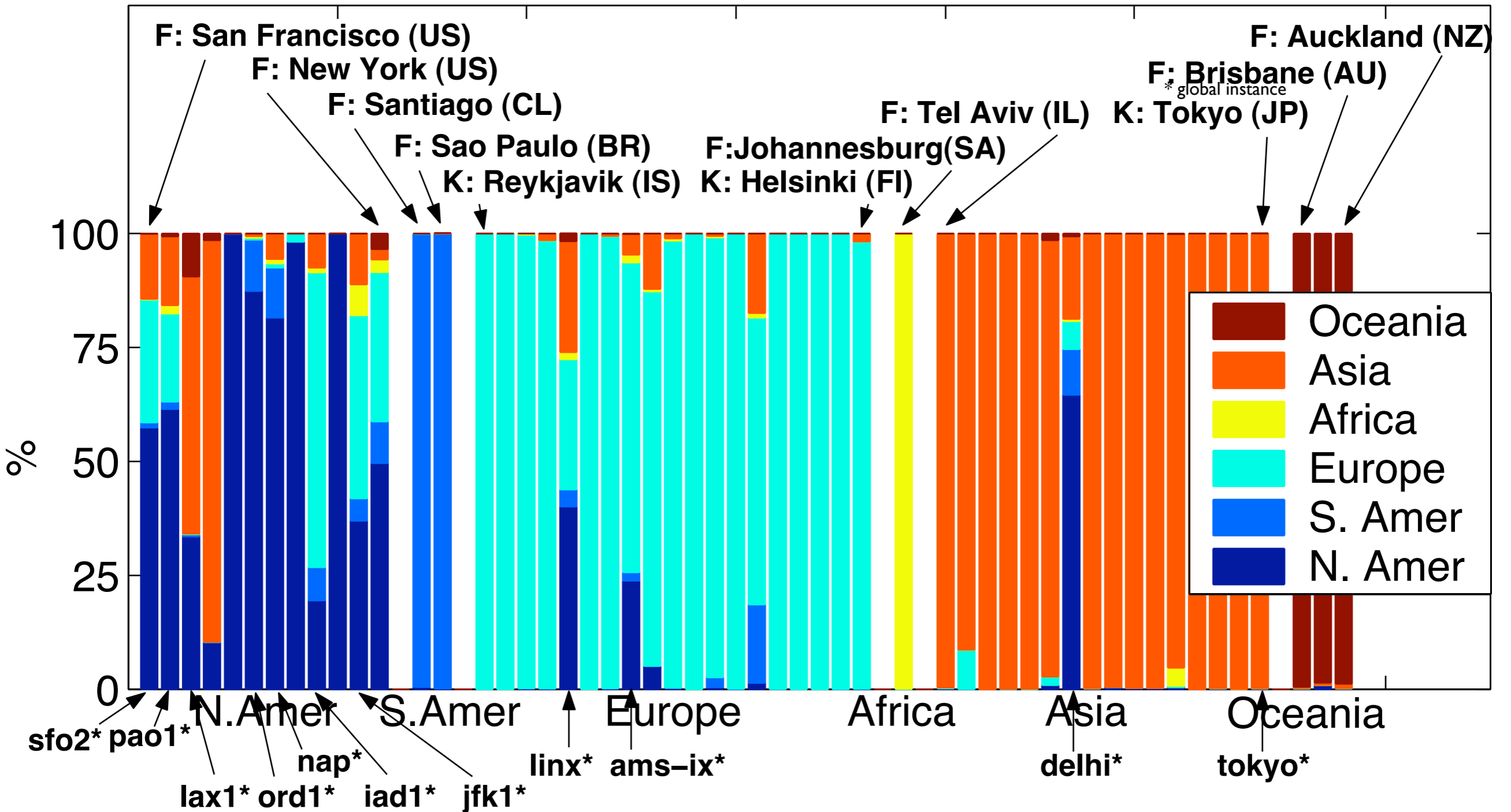


Instances sorted by the number of requests per second.



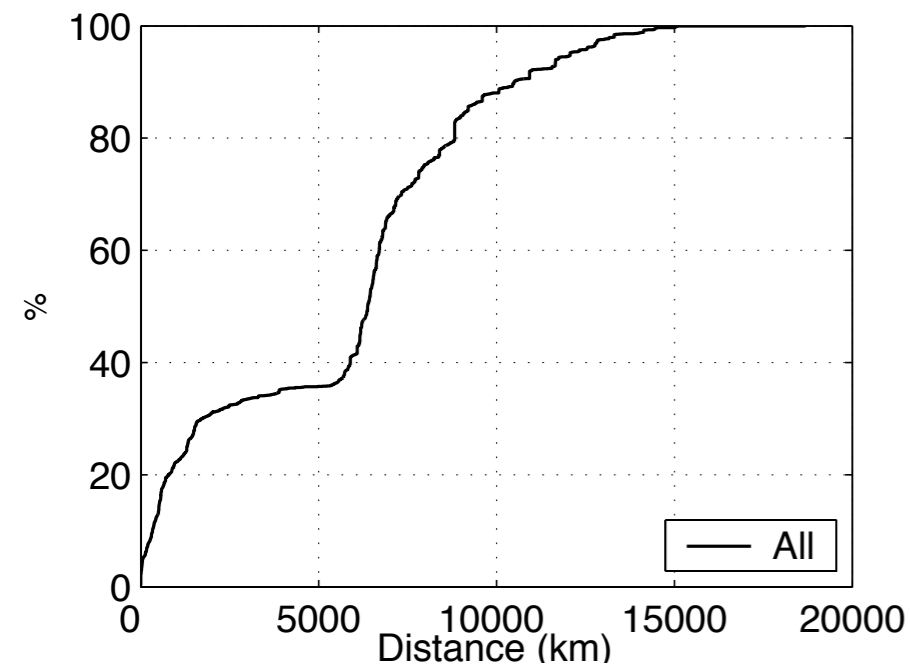
caida

Client Geographic locations

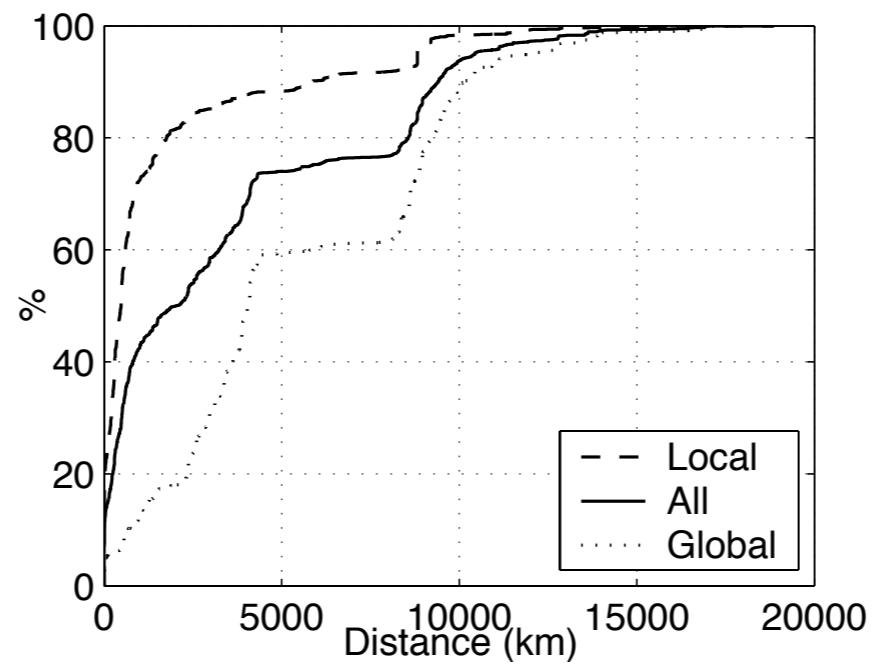




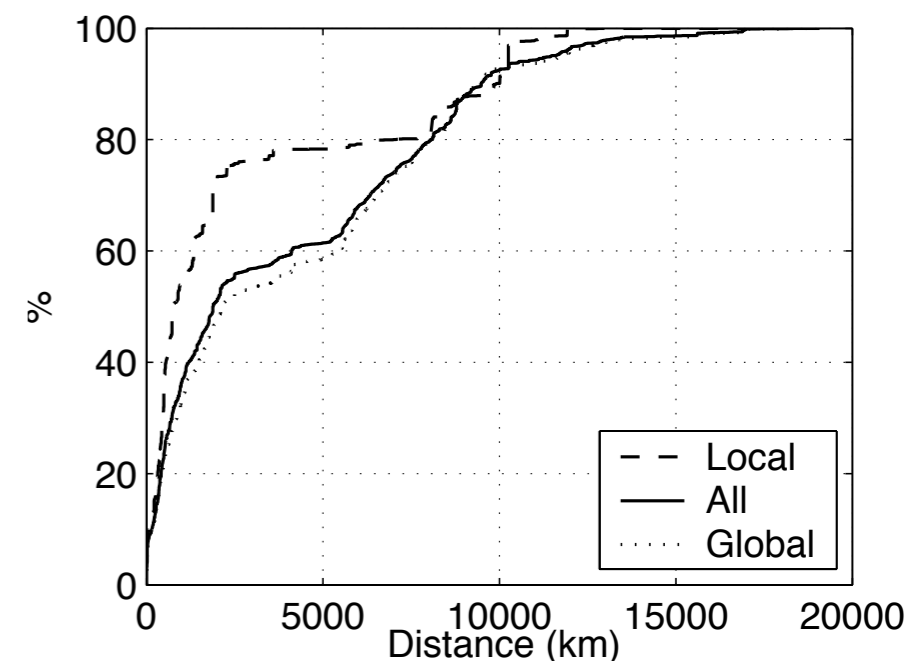
CDF of distance from instance to client



c-root



f-root

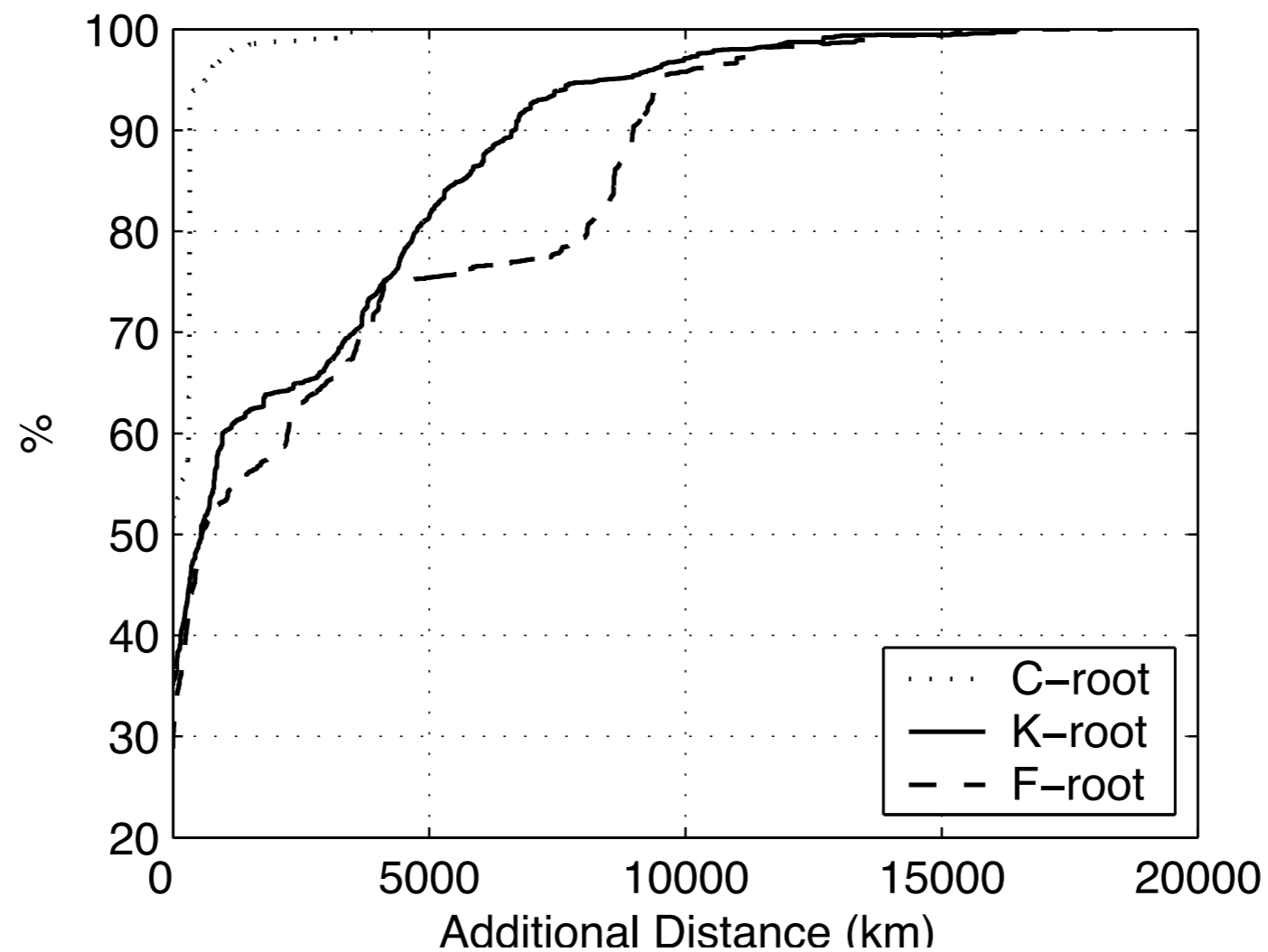


k-root

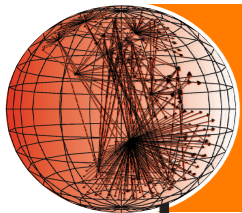
Inflection points between 5000 and 7000 km are the result of clients clustering on continents.



CDF of additional distance from optimal

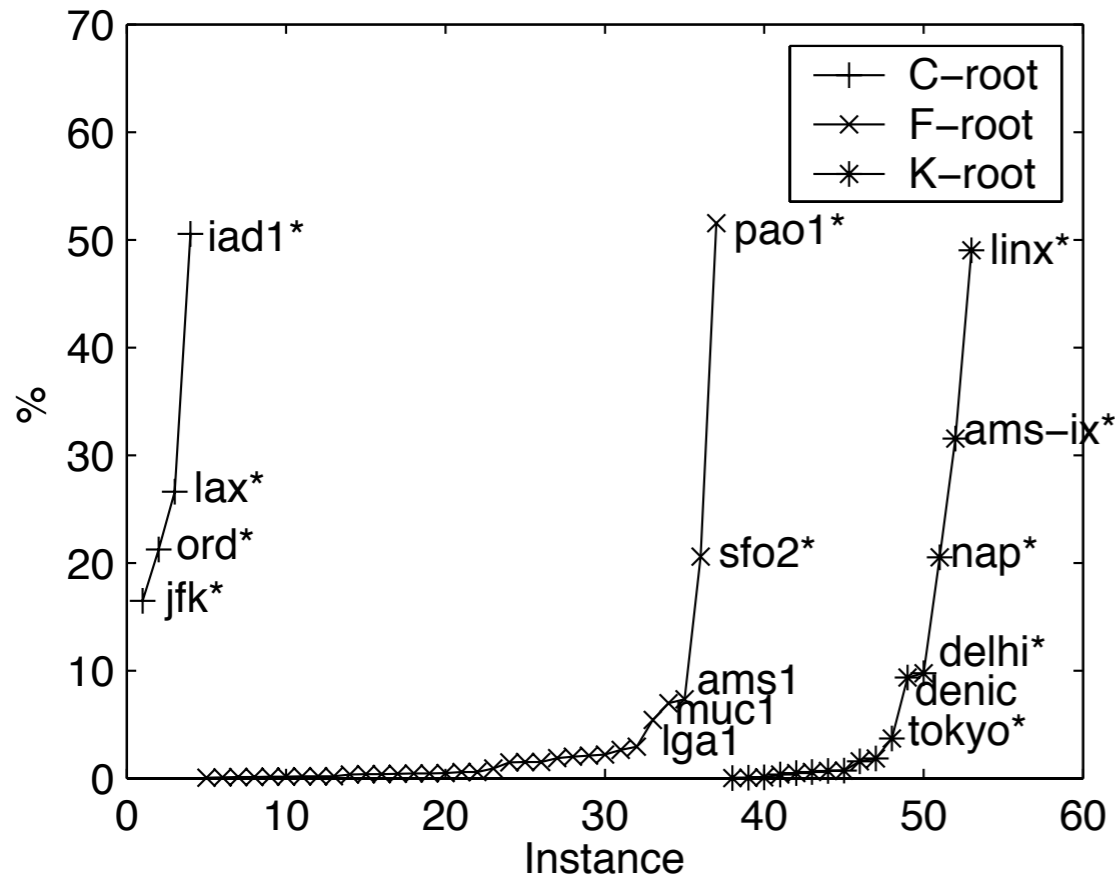


The additional distance a client's requests were routed from its geographically closest router.

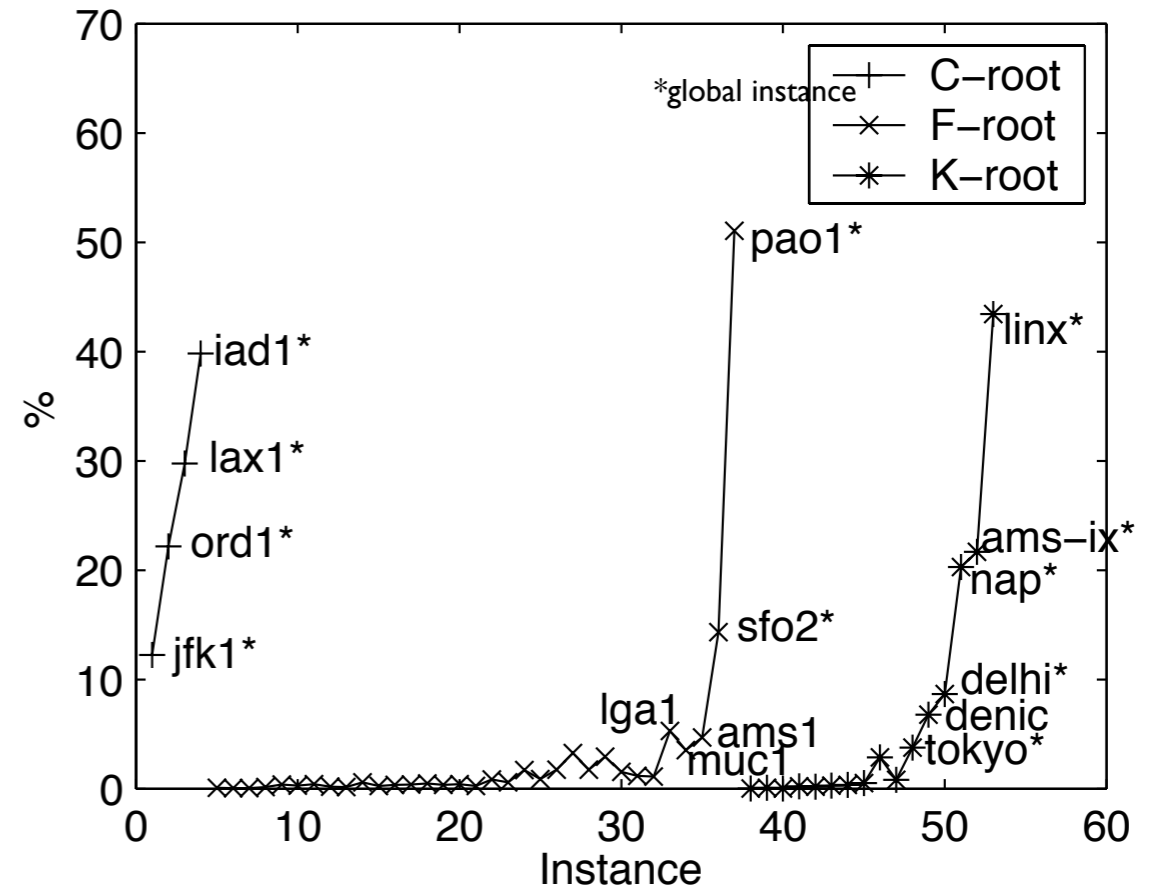


caida

Topological Coverage

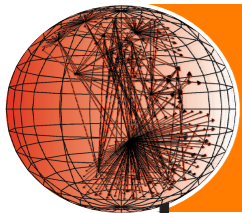


AS coverage



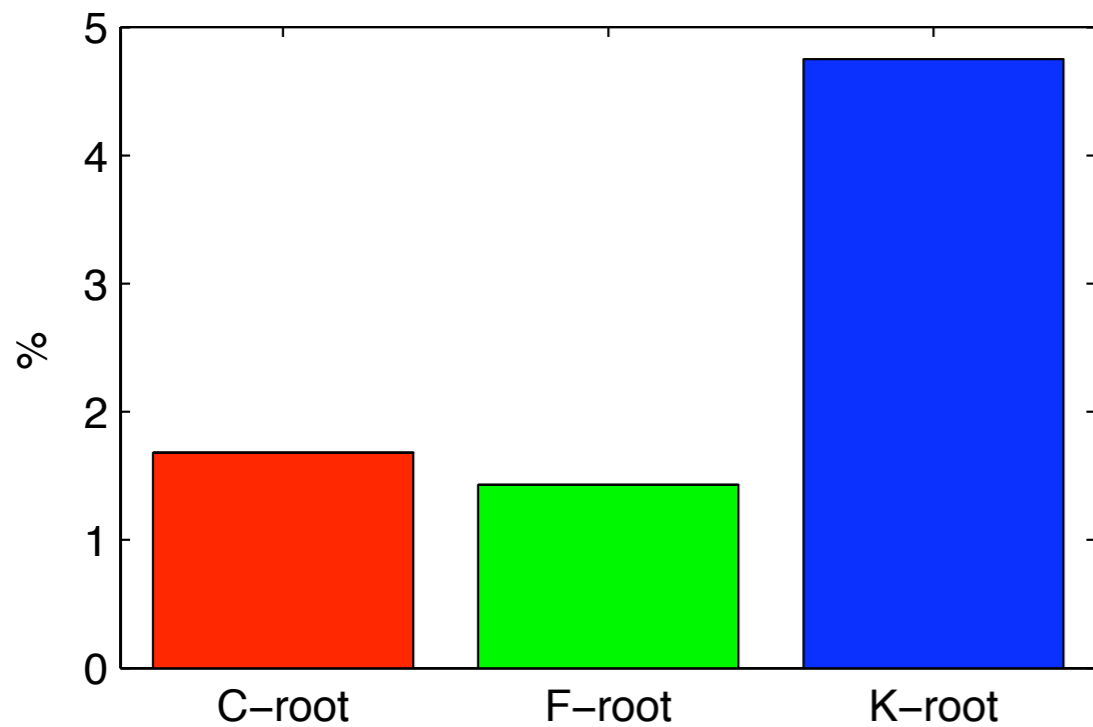
prefix coverage

Instances sorted by AS coverage.

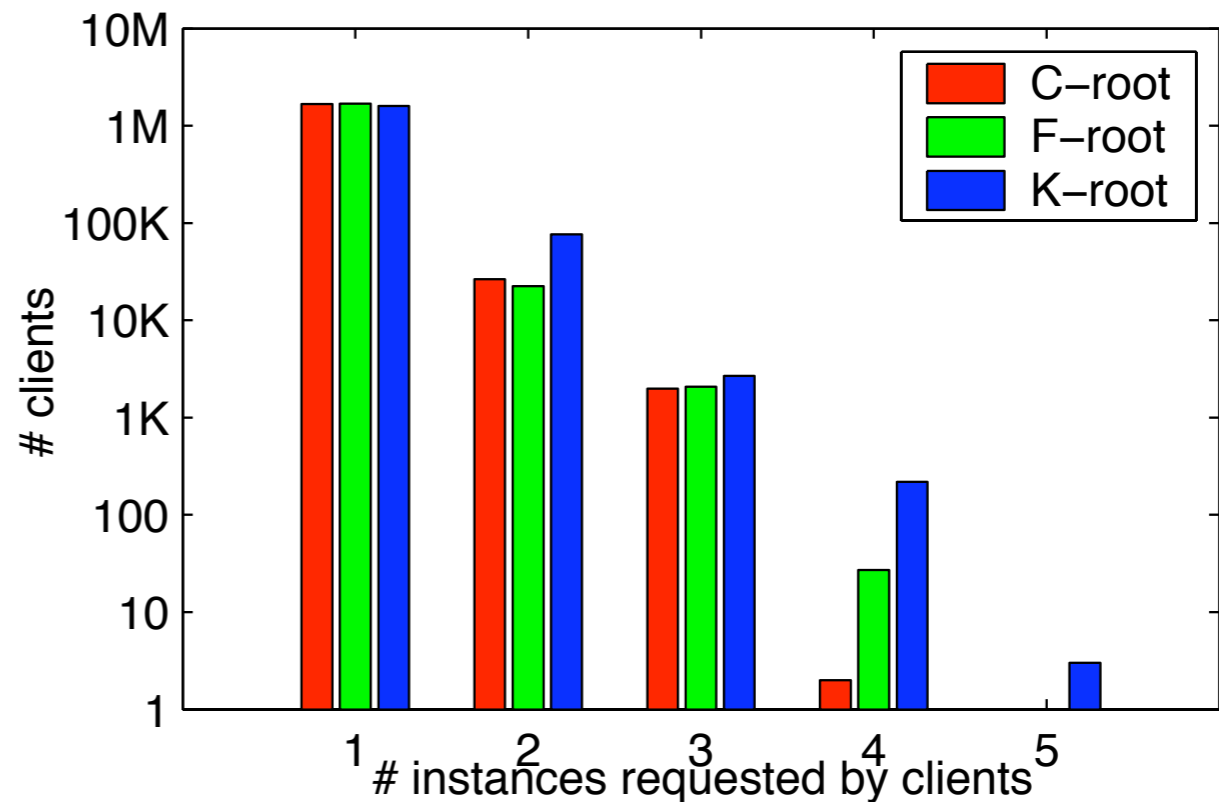


Client Stability

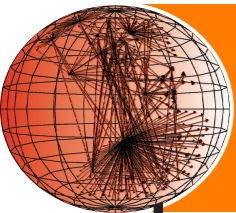
caida



the percentage of clients seen by multiple servers



the number of instances queried by clients



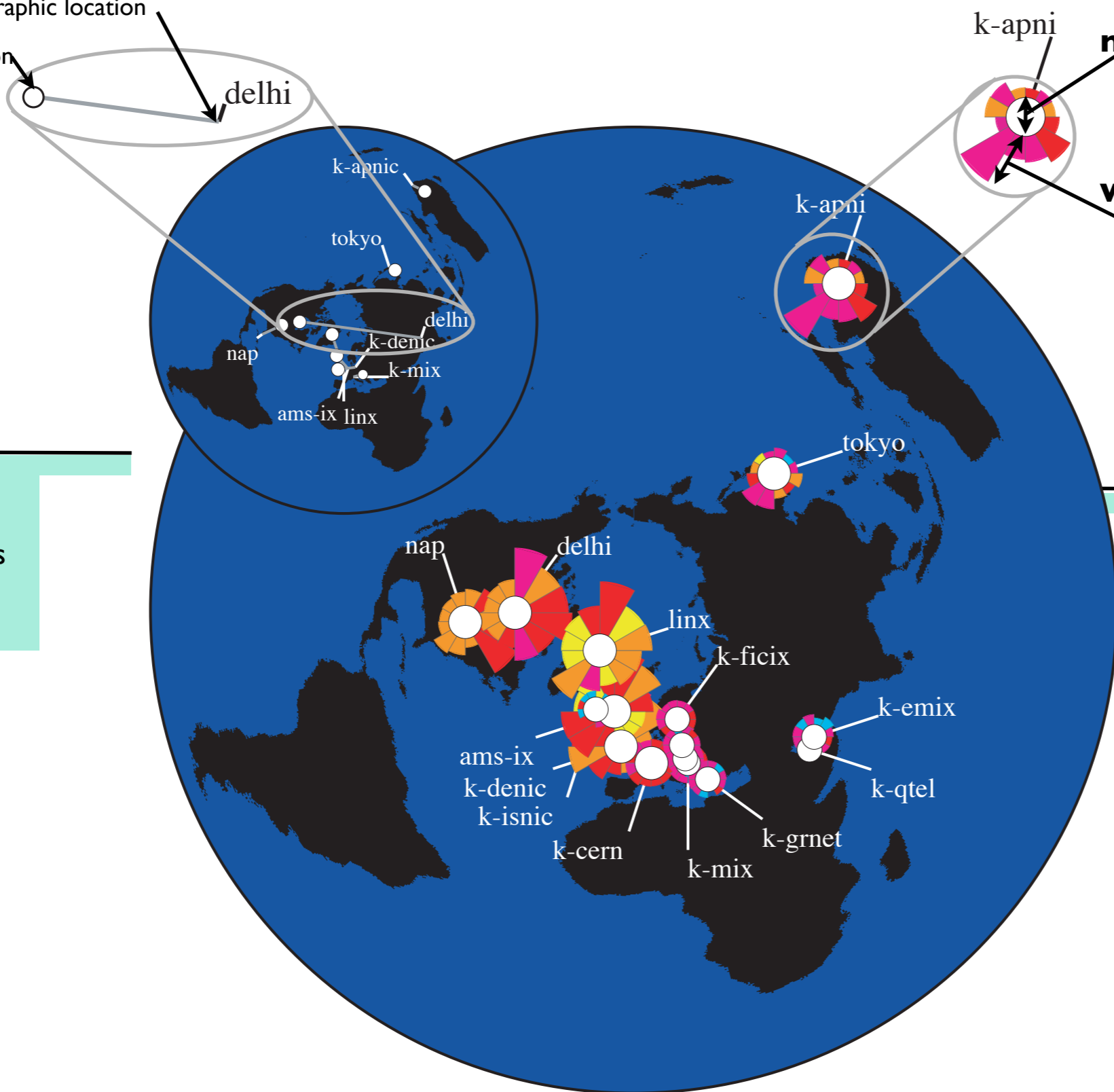
Influence Map Breakdown

caida

geographic location
displacement location

displacement

Distance from client's centroid and DNS server's geographic location.

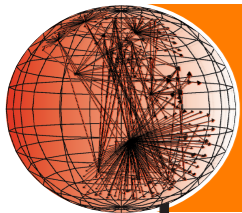


node
number of clients (size)

wedge
number of clients (color)
average distance to clients (size)

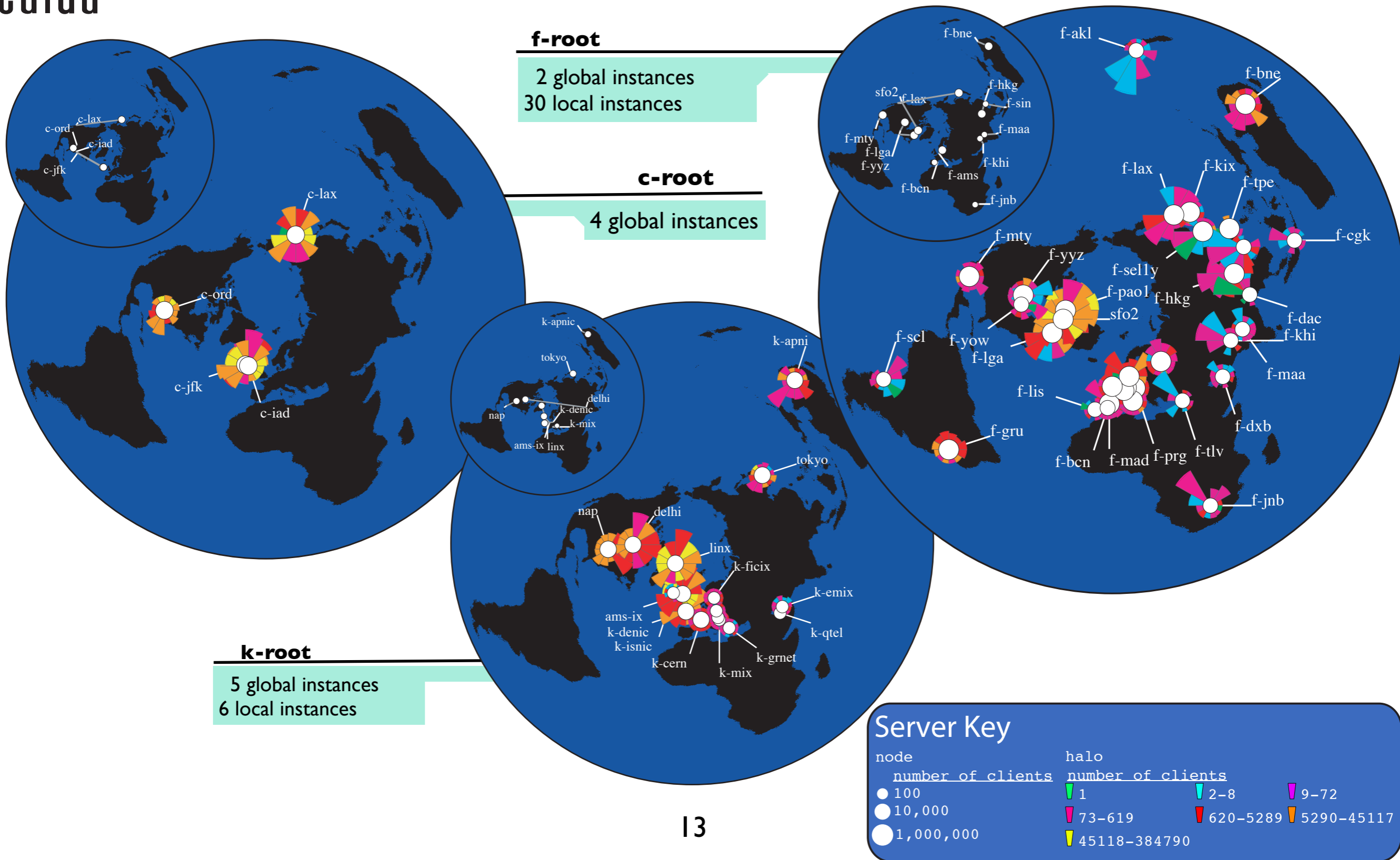
location

Number of DNS clients in a given direction and average distance in that direction.



caida

Influence Maps





conclusion

- Geographic clustering of clients to local instances is high.
 - 60% of clients experience small distance penalties between selected and optimal instances
 - local instances have strong diurnal patterns of use
- Small minority of clients experience a change in instance.
- ASes/IP addresses unevenly spread across instances, especially for f-root
- Propose second data collection 9th-10th, January 2007

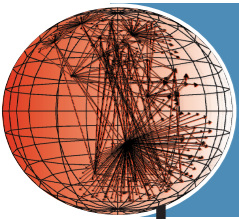


DatCat



<http://imdc.datcat.org>

 **DatCat** is an **NSF**
funded Internet measurement
data catalog.



caida

Target Problems



<http://imdc.datcat.org>

- data everywhere and lots of it
 - caida alone has over 50 terabytes of data
 - pcap (tcpdump) packet traces
 - skitter topology traces
 - routing tables
 - etc
- data is hard to find
 - no central indexing
 - many one-off data collections



Goals



<http://imdc.datcat.org>

Make the following processes easy for users.

- finding data sets of interest
 - Many researchers lack access and/or expertise to collect data needed for their research.
- adding new data sets to the catalog
 - Contributors (who are generally underfunded and providing data out of dedication to the general good) want to minimize time lost to their own research.
- annotating data sets in the catalog
 - Provides a flexible way for contributors and users to mark up interesting facts about data sets. Such as the number of packets or that a given file is corrupted.



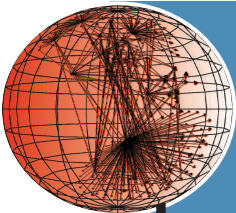
Current Status



<http://imdc.datcat.org>

current implementation

- user accounts
- data sets
 - only CAIDA data so far
 - 57,088 files indexed
 - 4.8 TB of data indexed
 - 11 separate collections
 - 100+ accounts registered
- browse simple/advanced search
- help/tutorial
- API for bulk contributions

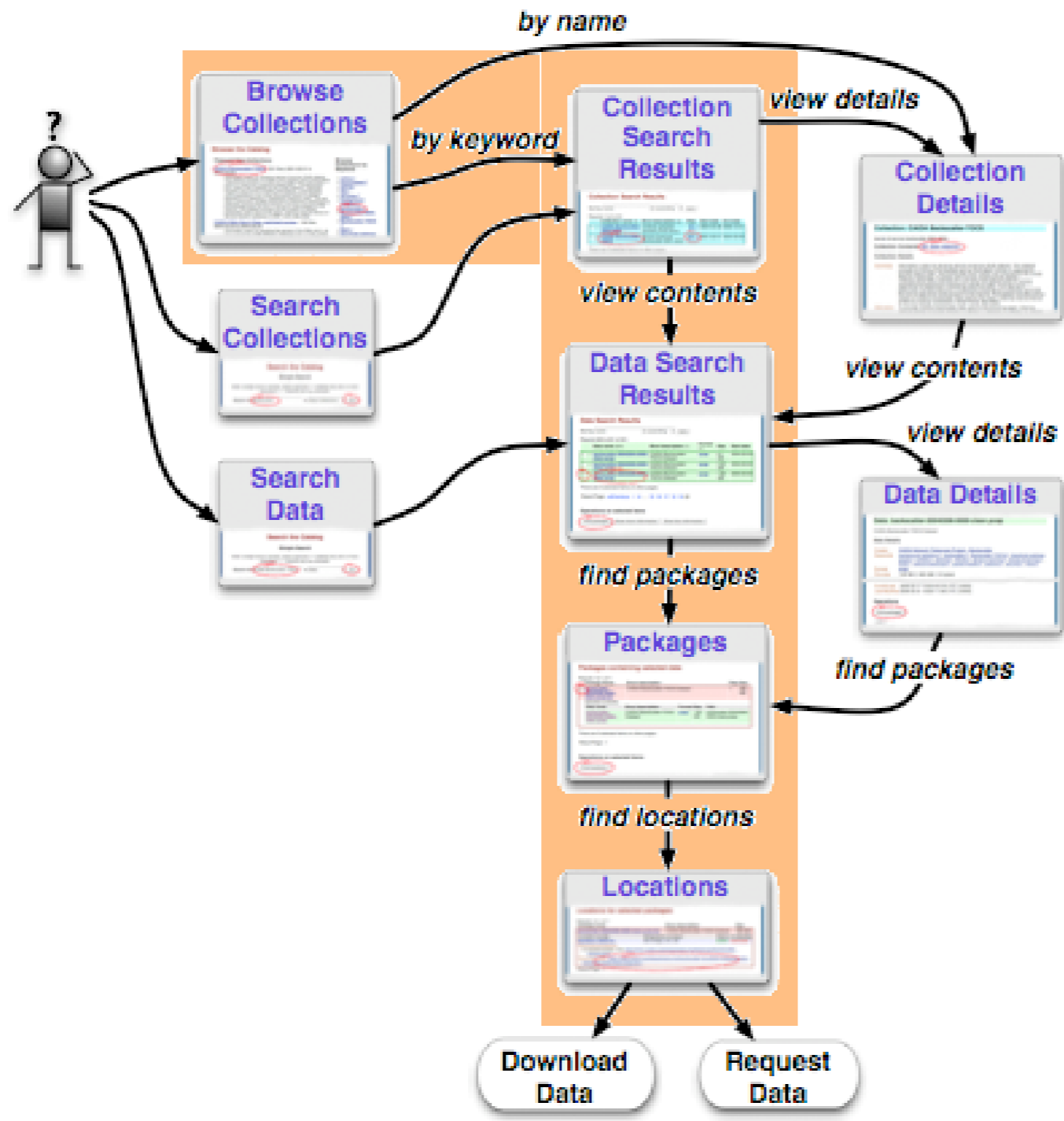


caida

Overview



<http://imdc.datcat.org>





Browse Catalog



<http://imdc.datcat.org>

Browse the Catalog

Featured Data Collections

[CAIDA Backscatter-TOCS](#) 231 files, 2001-02-01 to 2004-03-06
Information useful for studying denial-of-service (DoS) attacks. This dataset consists of 3 billion IPv4 packets sent by DoS attack victims in response to spoofed attack traffic. This backscatter from victims was collected by the UCSD Network Telescope. Possible uses include modeling DoS attacks, understanding victim populations, and using real packet traces to validate algorithms for detecting or classifying malicious traffic. This last use is particularly valuable because it is extremely challenging to artificially generate the kind of real-world noise present on the Internet. This dataset includes just the subset of CAIDA's backscatter data used for the paper "Inferring Internet Denial-of-Service Activity" published in ACM TOCS, May 2006.

[CAIDA Witty Worm Data, restricted access](#) - 132 files, 2004-03-20 to 2004-03-25
Information useful for studying the spread of the Witty worm, as observed by the UCSD Network Telescope over a 5-day period

Browse Collections by Keyword

- [active](#)
- [anonymized](#)
- [ARTS](#)
- [AS](#)
- [AS links](#)
- [background radiation](#)
- [backscatter](#)
- [Backscatter-2004-2005](#)
- [Backscatter-TOCS](#)
- [BGP](#)
- [blackhole address](#)

select collection of interest

- Browse page is a list of data file collections.
 - Collections are a high level group of related files.
 - A file may belong to more than one collection.
- User clicks on collection of interest to view the [collection details](#).



Collection Details



<http://imdc.datcat.org>

Collection: CAIDA Backscatter-TOCS

denial-of-service backscatter 2001-2004

Collection Contents: [231 data objects](#)

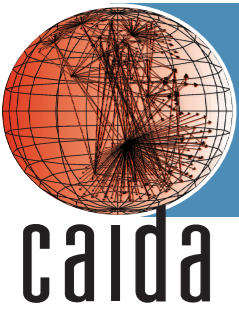
Collection Details

Summary Information useful for studying denial-of-service (DoS) attacks. This dataset consists of 3 billion IPv4 packets sent by DoS attack victims in response to spoofed attack traffic. This backscatter from victims was collected by the UCSD Network Telescope. Possible uses include modeling DoS attacks, understanding victim populations, and using real packet traces to validate algorithms for detecting or classifying malicious traffic. This last use is particularly valuable because it is extremely challenging to artificially generate the kind of real-world noise present on the Internet. This dataset includes just the subset of CAIDA's backscatter data used for the paper "Inferring Internet Denial-of-Service Activity" published in ACM TOCS, May 2006.

Motivation To provide CAIDA's backscatter data used for the following paper: Inferring Internet Denial-of-Service Activity. D. Moore, C. Shannon, D. Brown, G. Voelker

go to a list of collection's data objects

- Collection details displays specifics of the collection.
 - description of collection and its contents
 - motivation behind the collection
- If the collection matches the researchers' interest they can then [list the data objects](#).



Data Files



<http://imdc.datcat.org>

Data Search Results

Sort by

Results 229 to 231 of 231:

	Data name (231)	Short description (1)	Format (1)	Size	Start date
<input type="checkbox"/>	backscatter-20040304-0000-clean.pcap	CAIDA Backscatter-TOCS Dataset	pcap	8.1 GB	2004-03-03
<input type="checkbox"/>	backscatter-20040305-0000-clean.pcap	CAIDA Backscatter-TOCS Dataset	pcap	5.32 GB	2004-03-04
<input checked="" type="checkbox"/>	backscatter-20040306-0000-clean.pcap	CAIDA Backscatter-TOCS Dataset	pcap	1.82 GB	2004-03-05

There are 0 selected items on other pages.

Result Page: [<<Previous](#) [1](#) [10](#) ... [15](#) [16](#) [17](#) [18](#) [19](#) [20](#)

Operations on selected items

select desired data

find the data's packages.

- A listing of the data files contained within the selected collection. Also shown: general statistics about the files such as format and size.
- After the user has selected the set of files they are interested in, they then **find packages** which contain them.

Packages containing selected data

Results 1 to 1 of 1:

Package Name	Short description	Files	Size
<input checked="" type="checkbox"/> backscatter-20040306-0000-clean.pcap.lzo	CAIDA Backscatter-TOCS Dataset	1	566 MB

Selected contents:

Data name	Short description	Format	Size	Path
backscatter-20040306-0000-clean.pcap	CAIDA Backscatter-TOCS Dataset	pcap	1.82 GB	backscatter-20040306-0000-clean.pcap

There are 0 selected items on other pages.

Result Page: 1

Operations on selected items

select desired packages

find download locations

- Packages are the downloadable groupings of one or more data files.
- Once the user has selected a set of packages, he then **finds download locations**.

Locations for selected packages

Results 1 to 1 of 1:

Package name	Short description	Size
backscatter-20040306-0000-clean.pcap.izo	CAIDA Backscatter-TOCS Dataset	566 MB

Location handle	Geographic location	Status	Availability
/location/1-3NCP-G	San Diego, CA, US	active	restricted

Download procedure:
to request access, visit: http://www.caida.org/analysis/security/telescope/backscatter_request.xml

Download URL: <https://data.caida.org/datasets/security/backscatter-tocs/2004-03/backscatter-20040306-0000-clean.pcap.izo>

Result Page: 1

select a location for the package

- Locations are the place or process by which the package can be obtained.
 - Provides a simple URL or instructions which must be followed to get the data.
- Get your data! 😊



Future Work



<http://imdc.datcat.org>

- small number of invited third party contributors
- public contributors
- tools
- studies
 - collections specialized for papers / experiments / etc