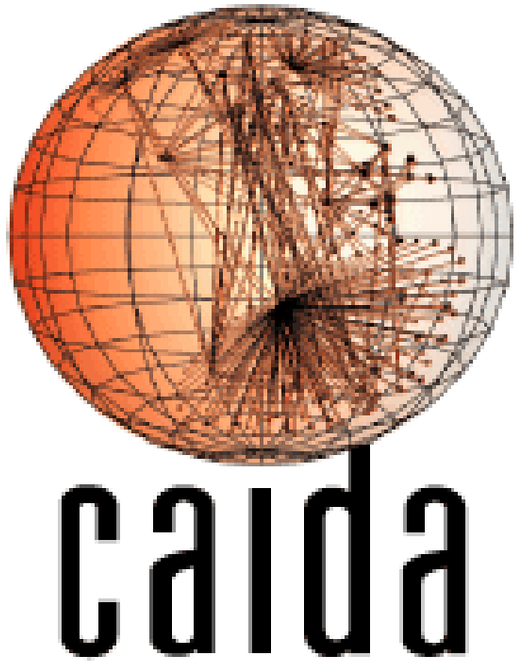


Two days in the life of three DNS root servers + DatCat[SM] overview and demonstration

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 toplogical clustering of DNS clients.

Data collected by OARC ISC with COGENT and RIPE NCC.

Outline

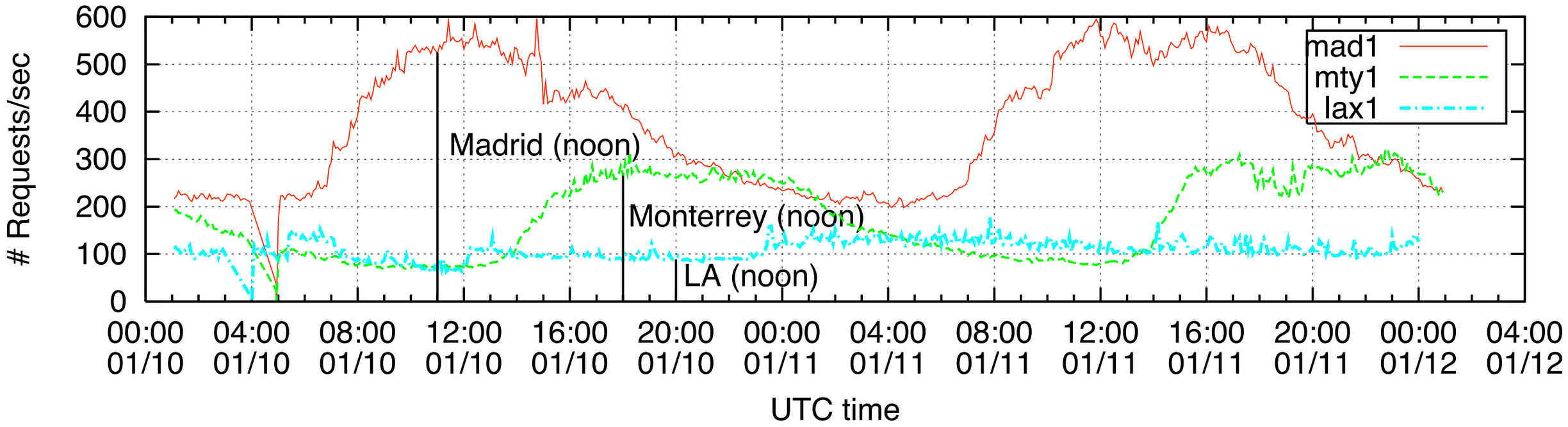
DNS anycast analysis

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

Data Source

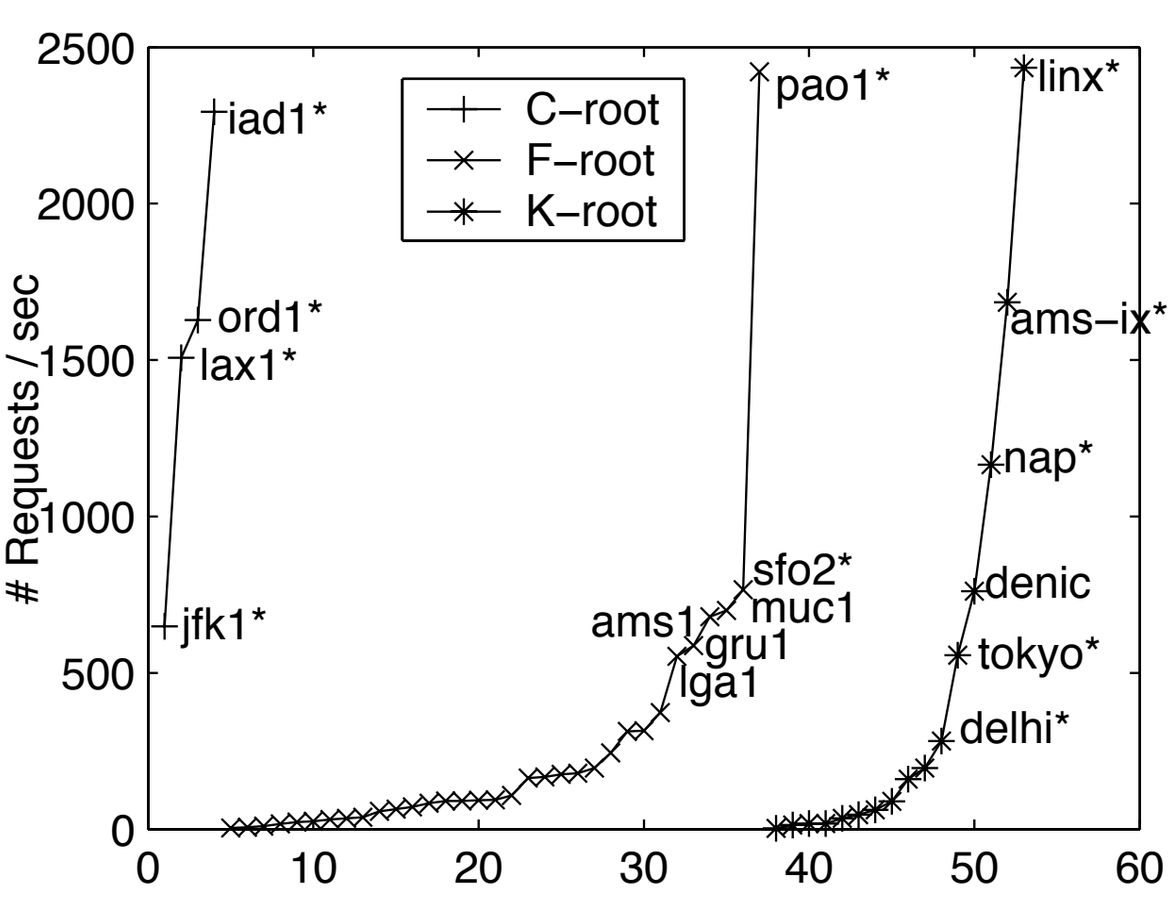
- dates
  - January 10th-11th 2006 (47 hours)
- DNS sources
  - c-root (Cogent) 4 instances
  - f-root (ISC) 32 instances
  - k-root (RIPE) 11 instances
- geographic
  - Netacuity database used for geographic mapping
- topological
  - Route Views used for ASs and prefixes (January 10th)

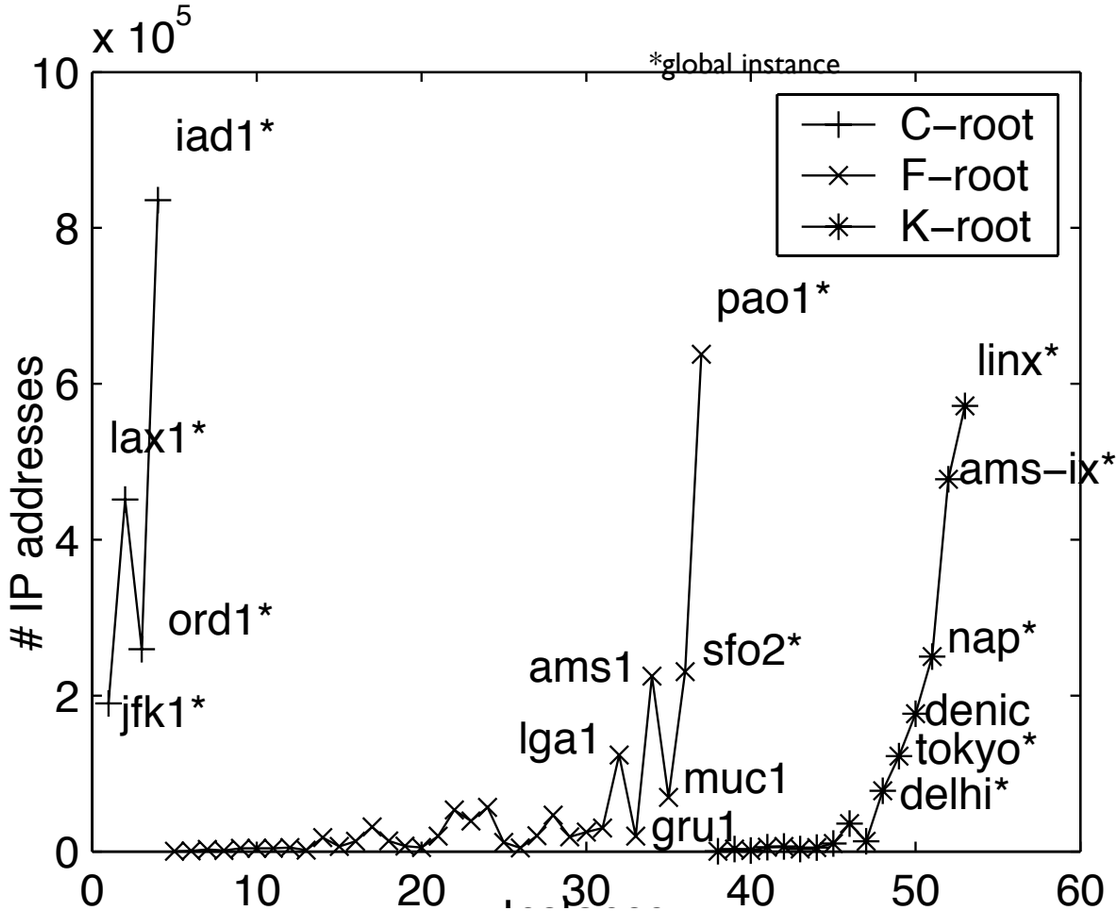
Diurnal Patterns



The local instances mad1 and mty1 show clear diurnal patterns

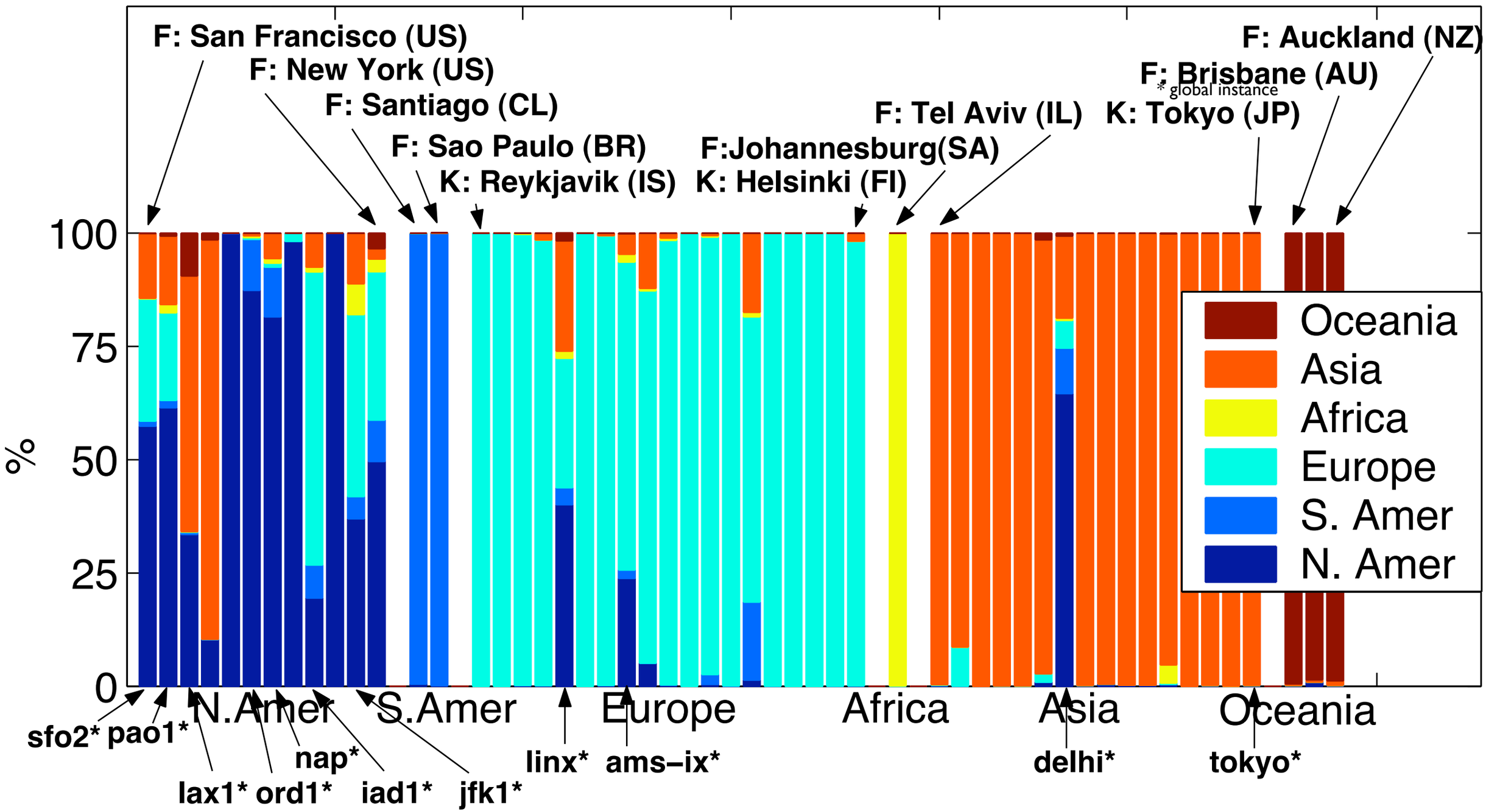
Num. of Requests/Addresses



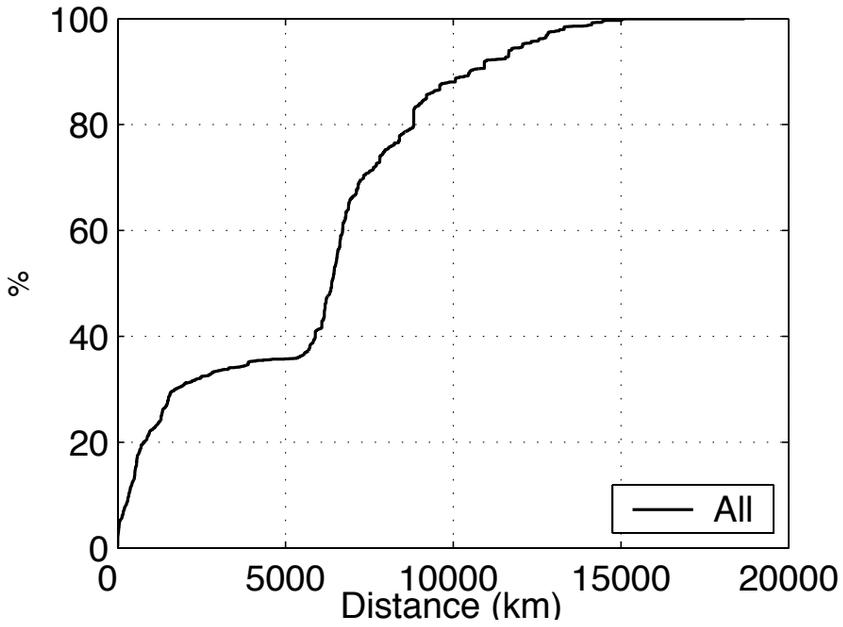


Instances sorted by the number of requests per second.

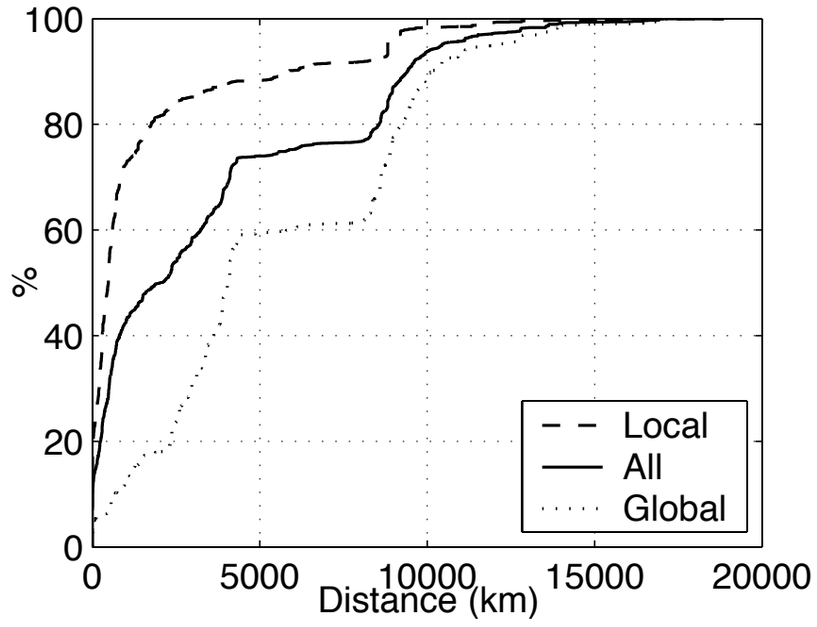
Client Geographic locations



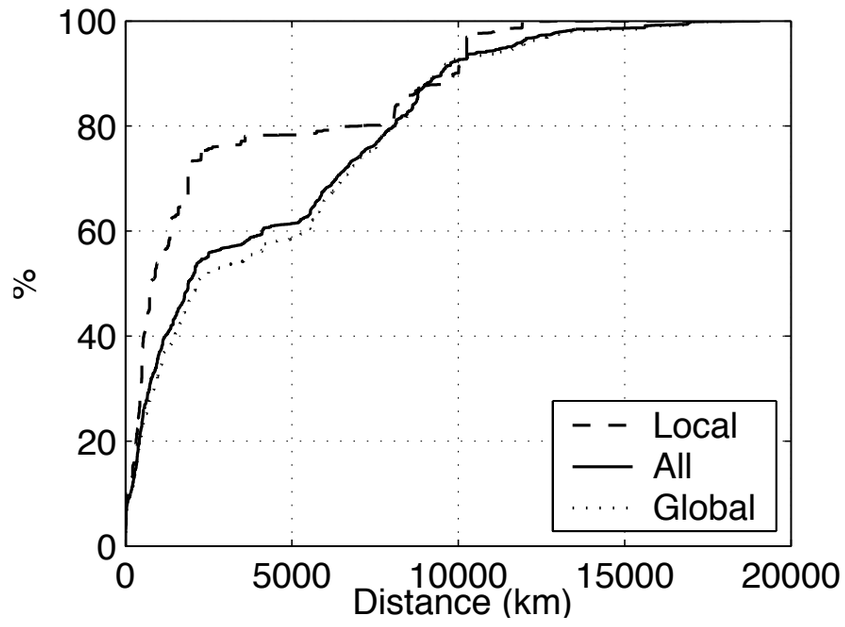
CDF of distance from instance to client



c-root



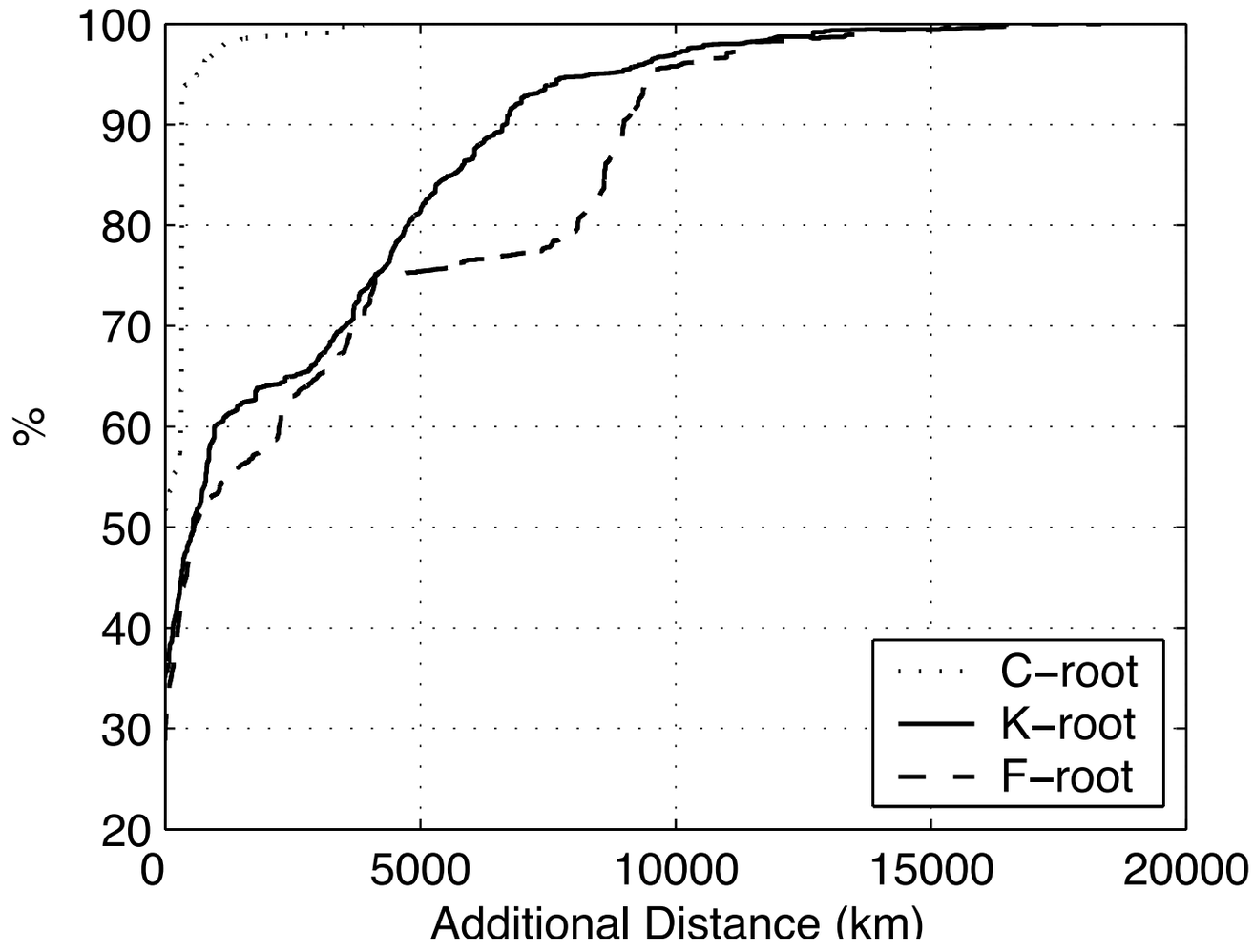
f-root



k-root

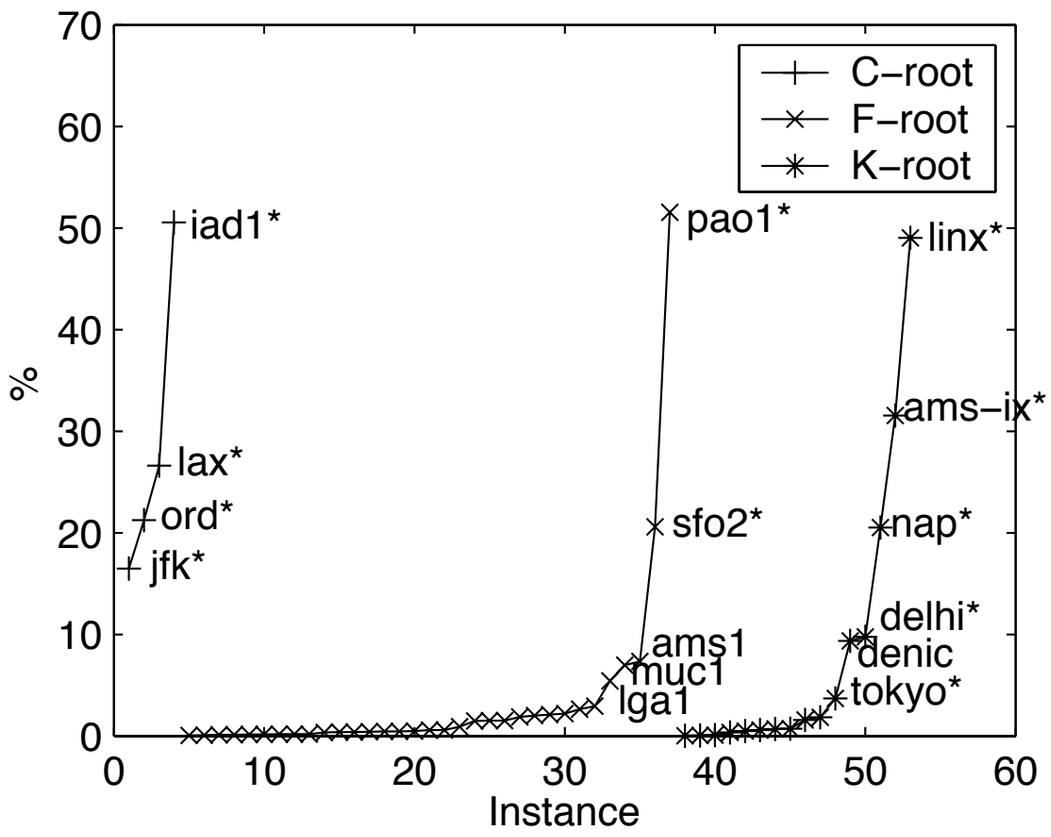
Inflection points between 5000 and 7000 km are the result of clients clustering on continents.

CDF of additional distance from optimal

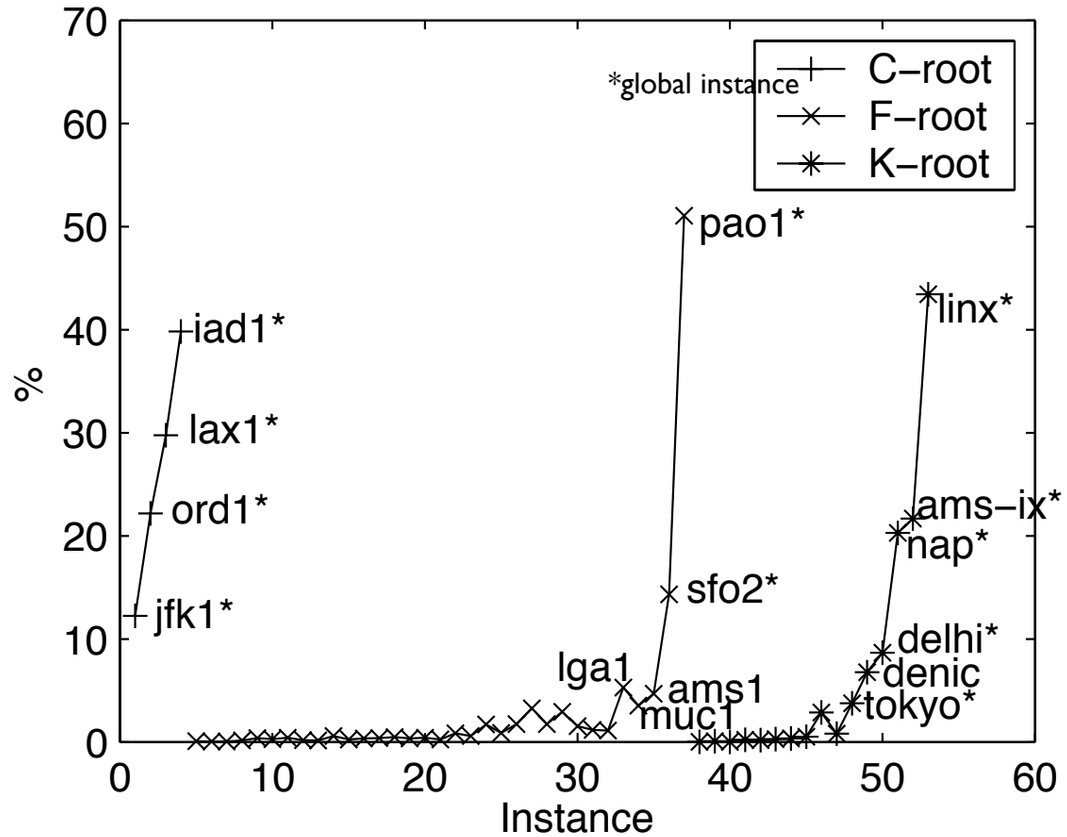


The additional distance a client's requests were routed from its geographically closest router.

Topological Coverage



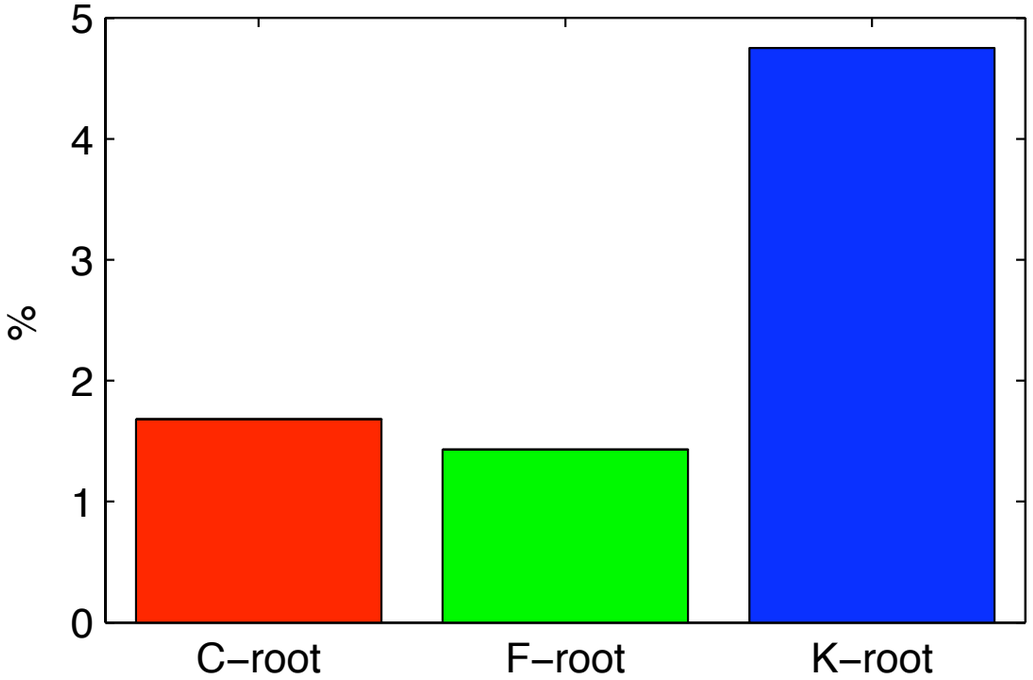
AS coverage



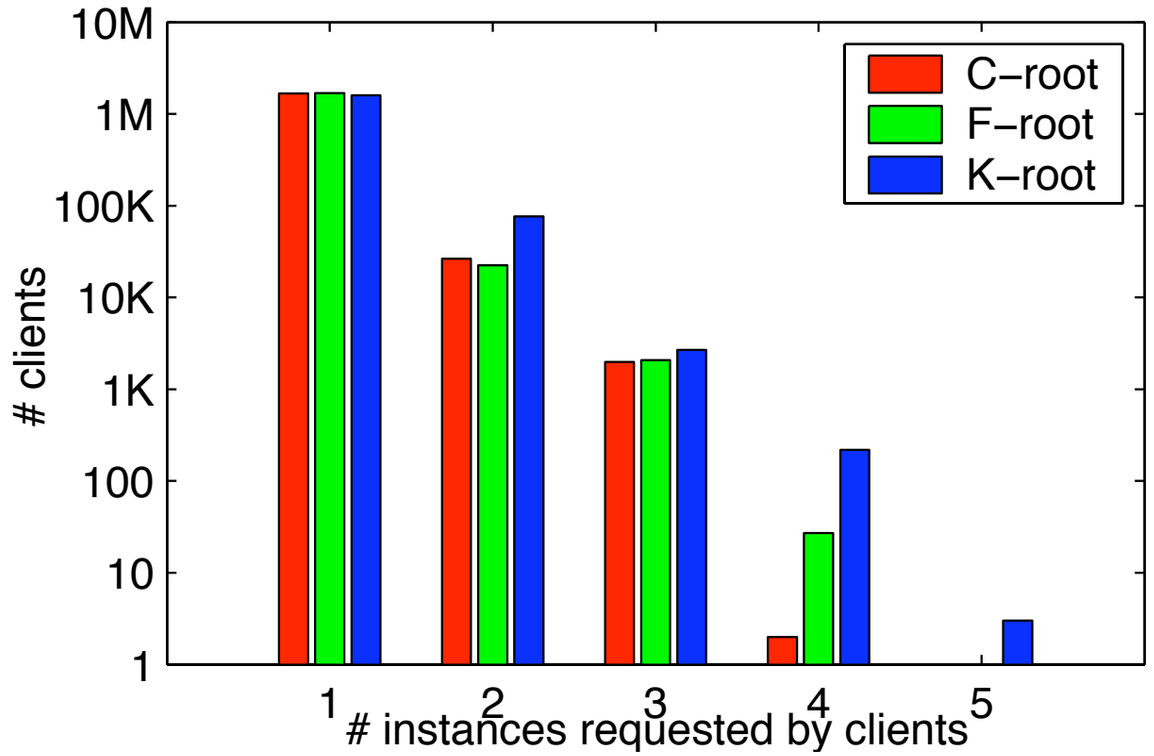
prefix coverage

Instances sorted by AS coverage.

Client Stability



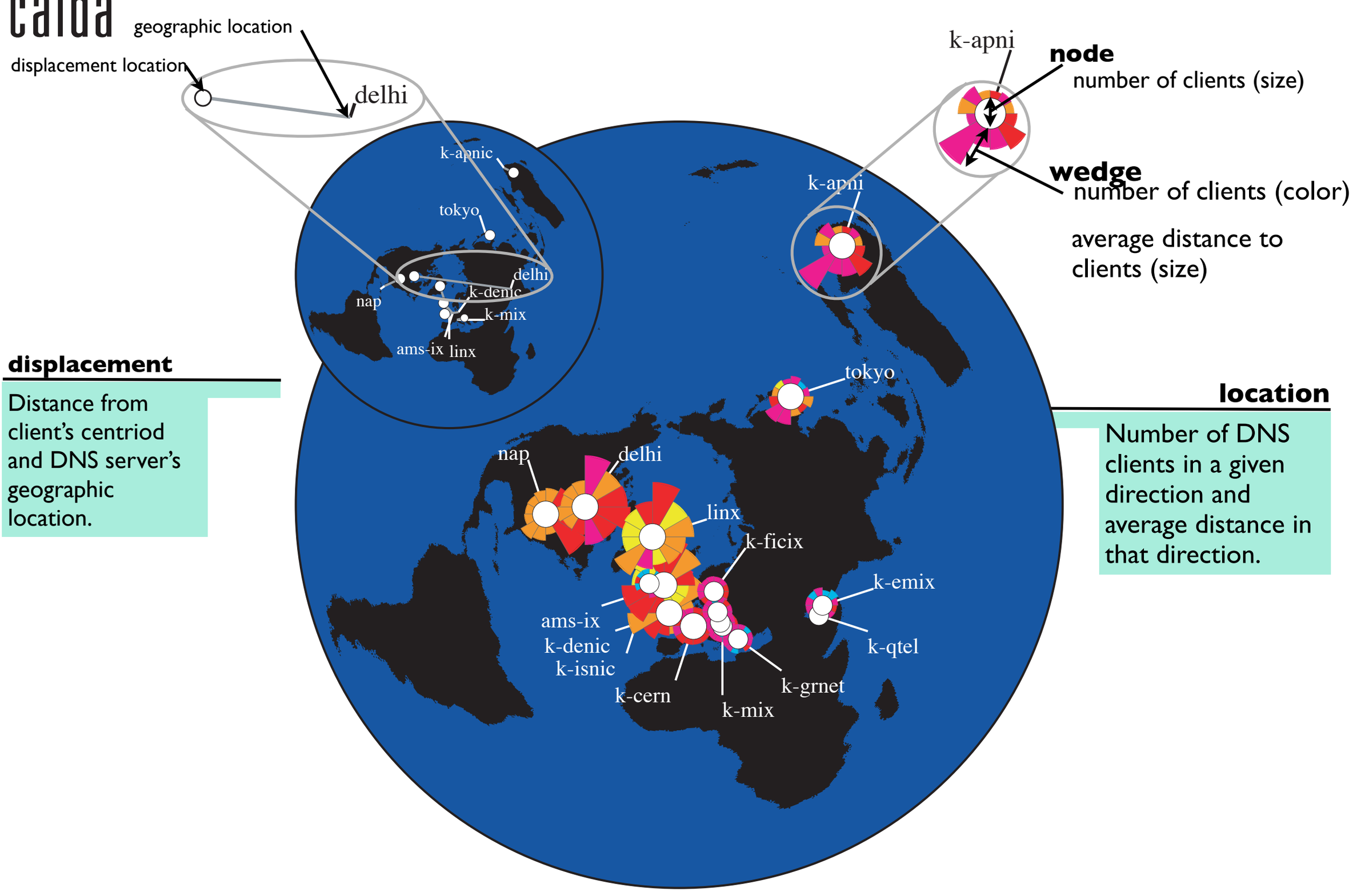
the percentage of clients seen by multiple servers



the number of instances queried by clients

Influence Map Breakdown

caida



displacement

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

location

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

Influence Maps

caida

f-root

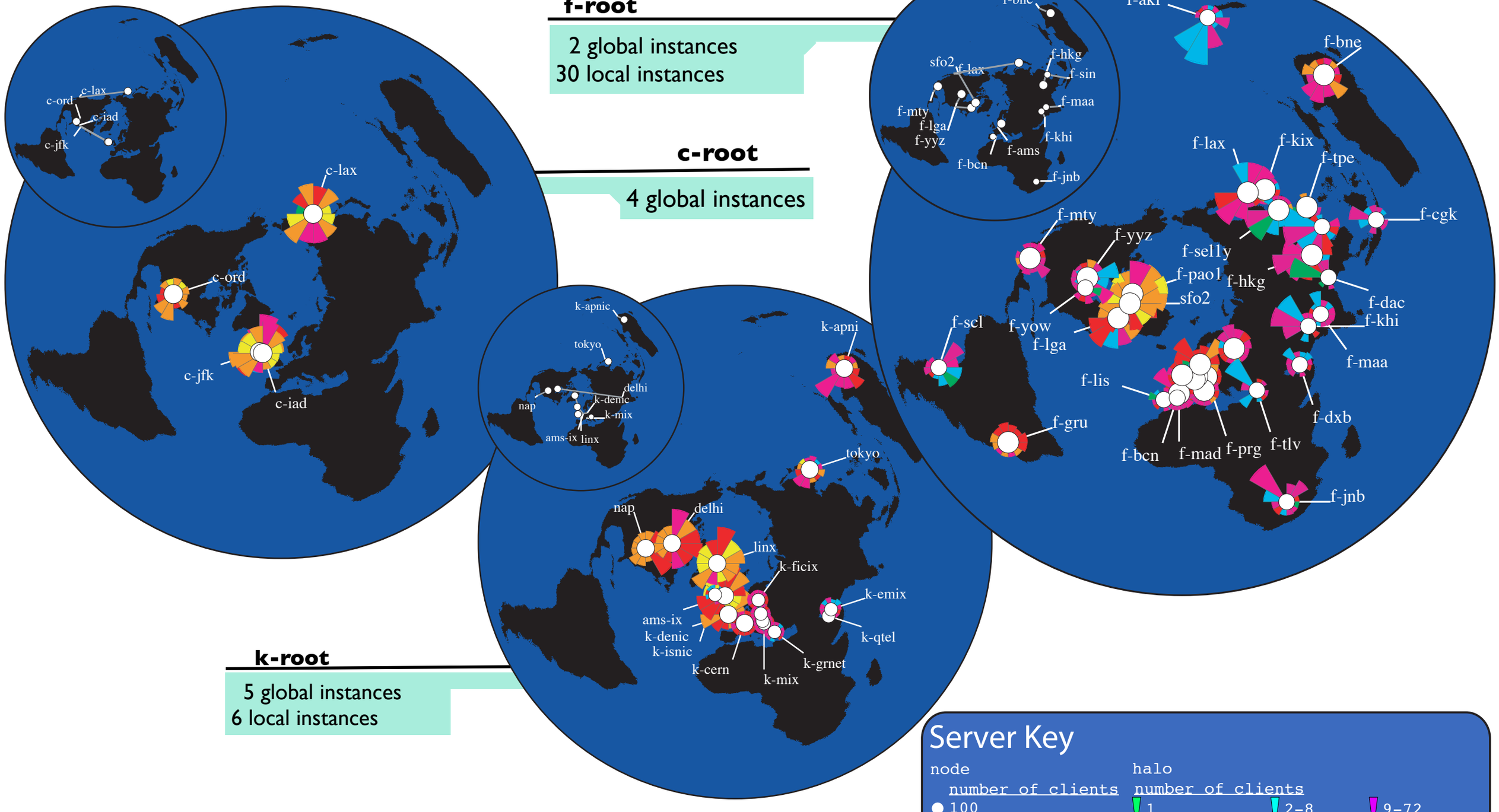
2 global instances
30 local instances

c-root

4 global instances

k-root

5 global instances
6 local instances



Server Key

node
number of clients
100
10,000
1,000,000

halo
number of clients
1
2-8
9-72
73-619
620-5289
5290-45117
45118-384790

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 colletion 9th-10th, January 2007

DatCat is an NSF funded Internet measurement data catalog.

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

caida

DatCat

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

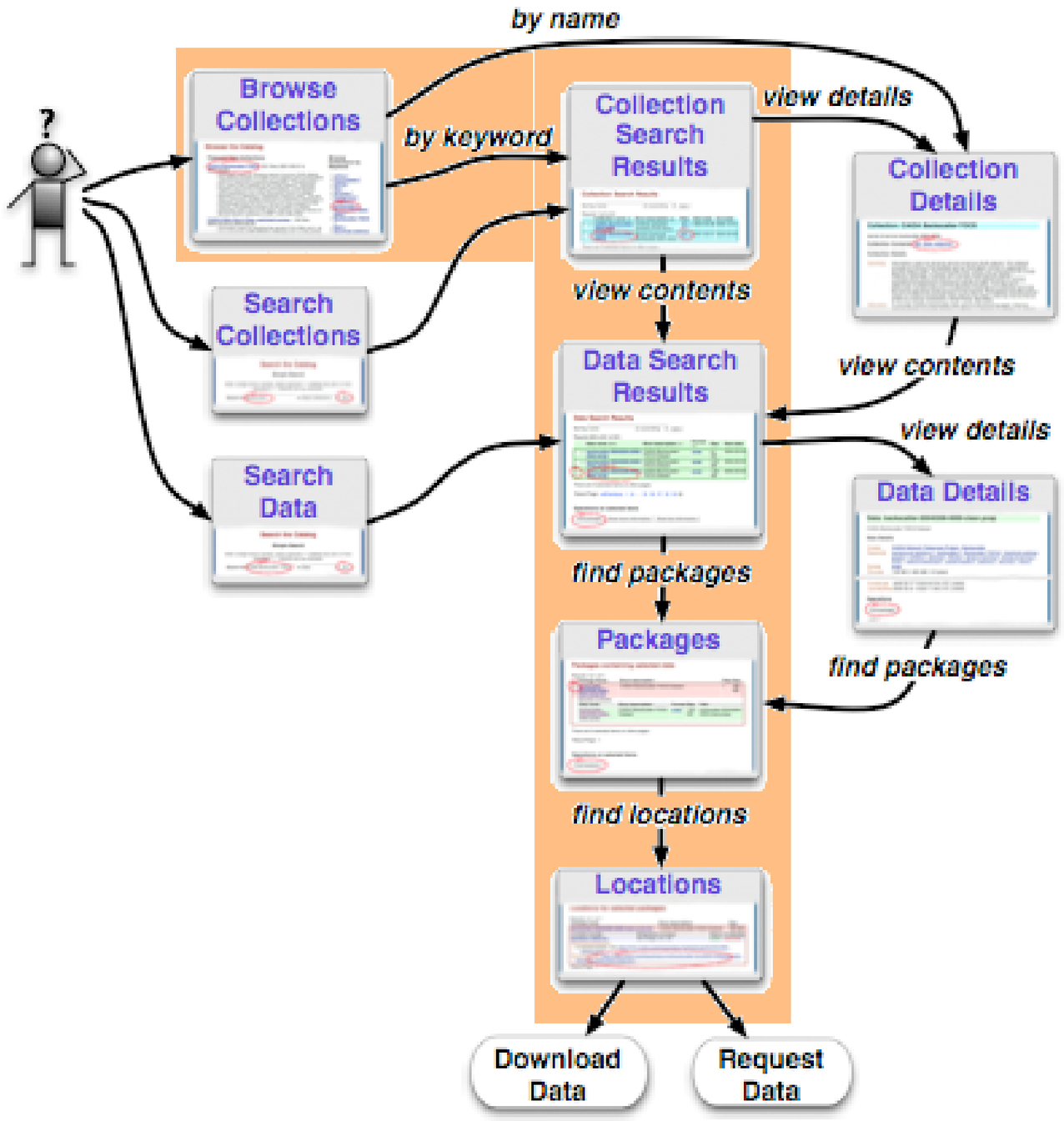
http://imdc.datcat.org

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

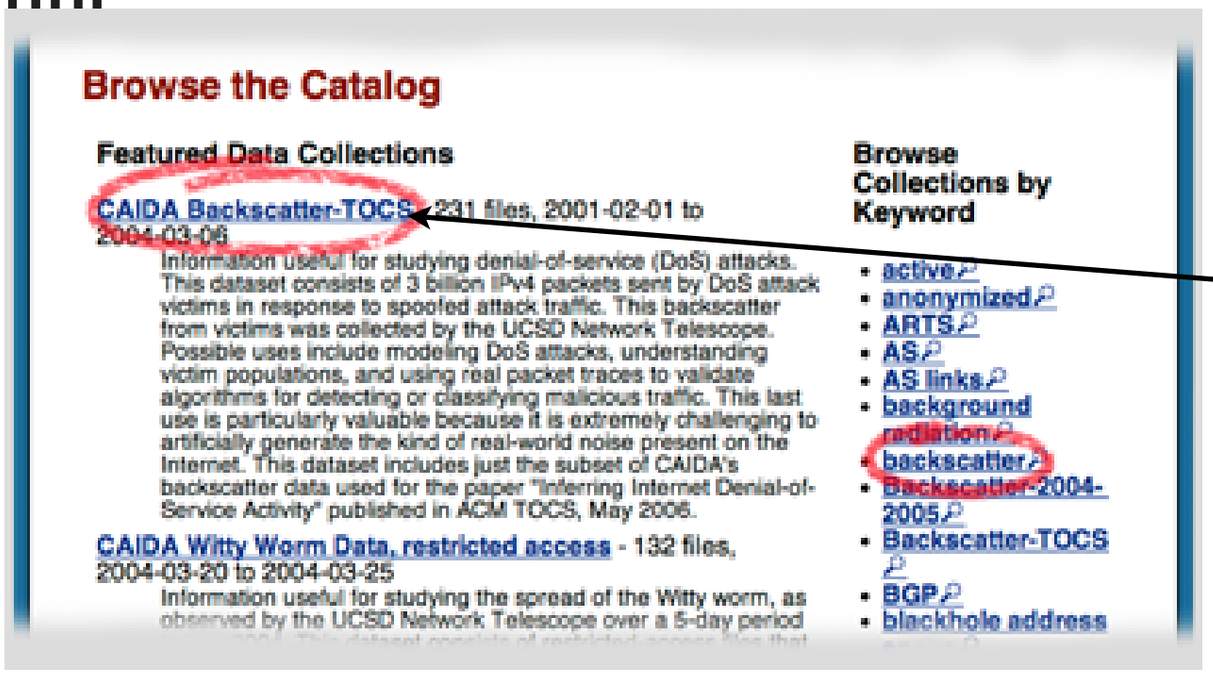
Overview

http://imdc.datcat.org



Browse Catalog

http://imdc.datcat.org

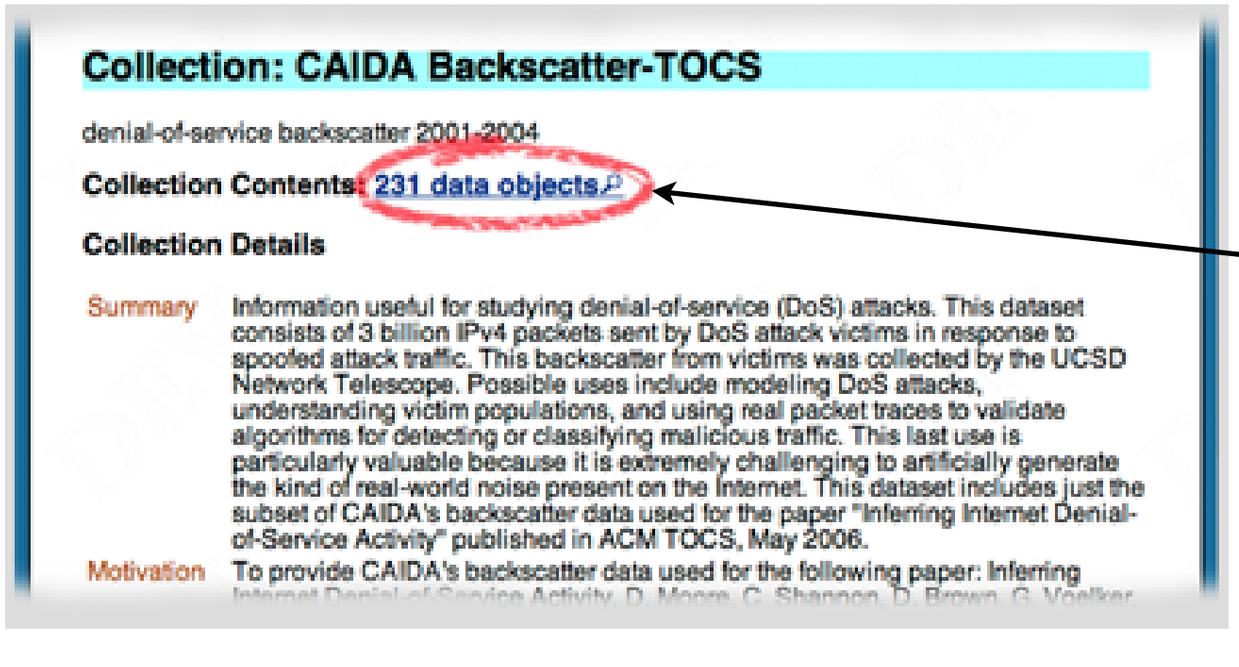


select collection of interest

- 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

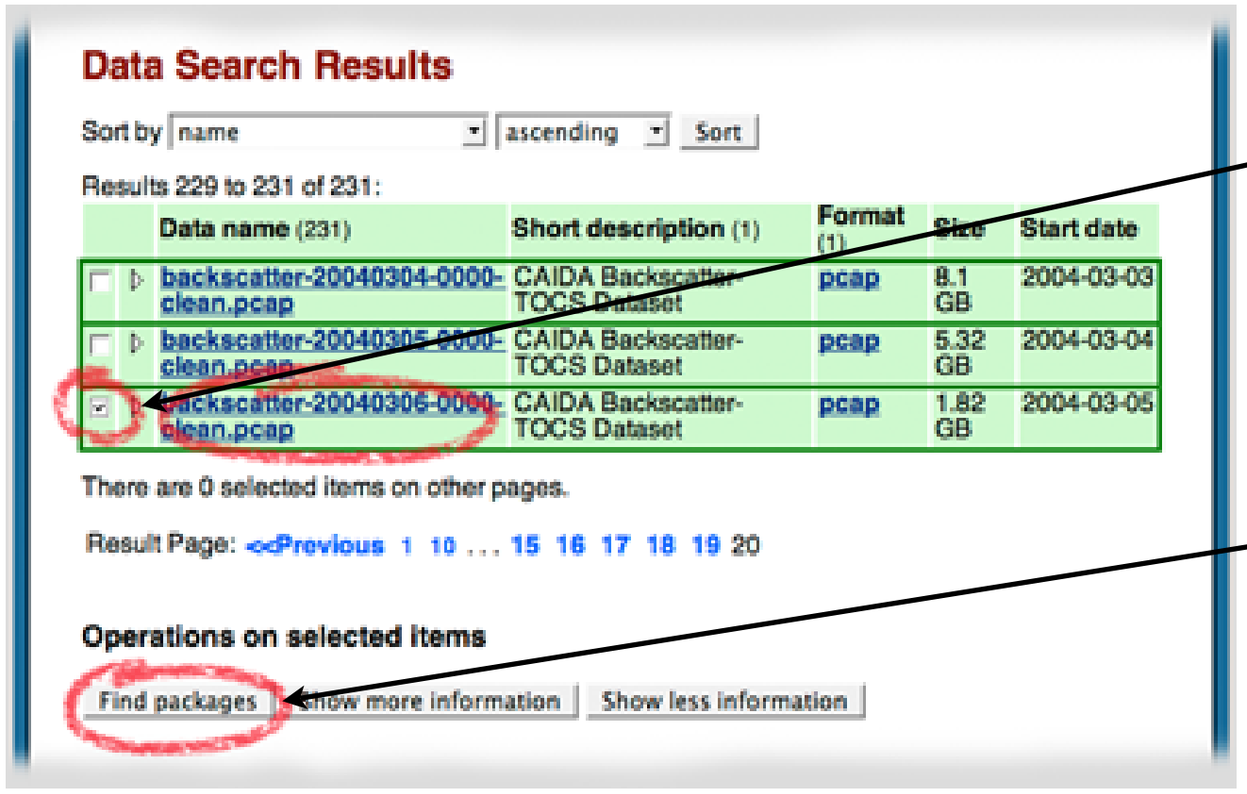


go to a list of collection's data objects

- 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



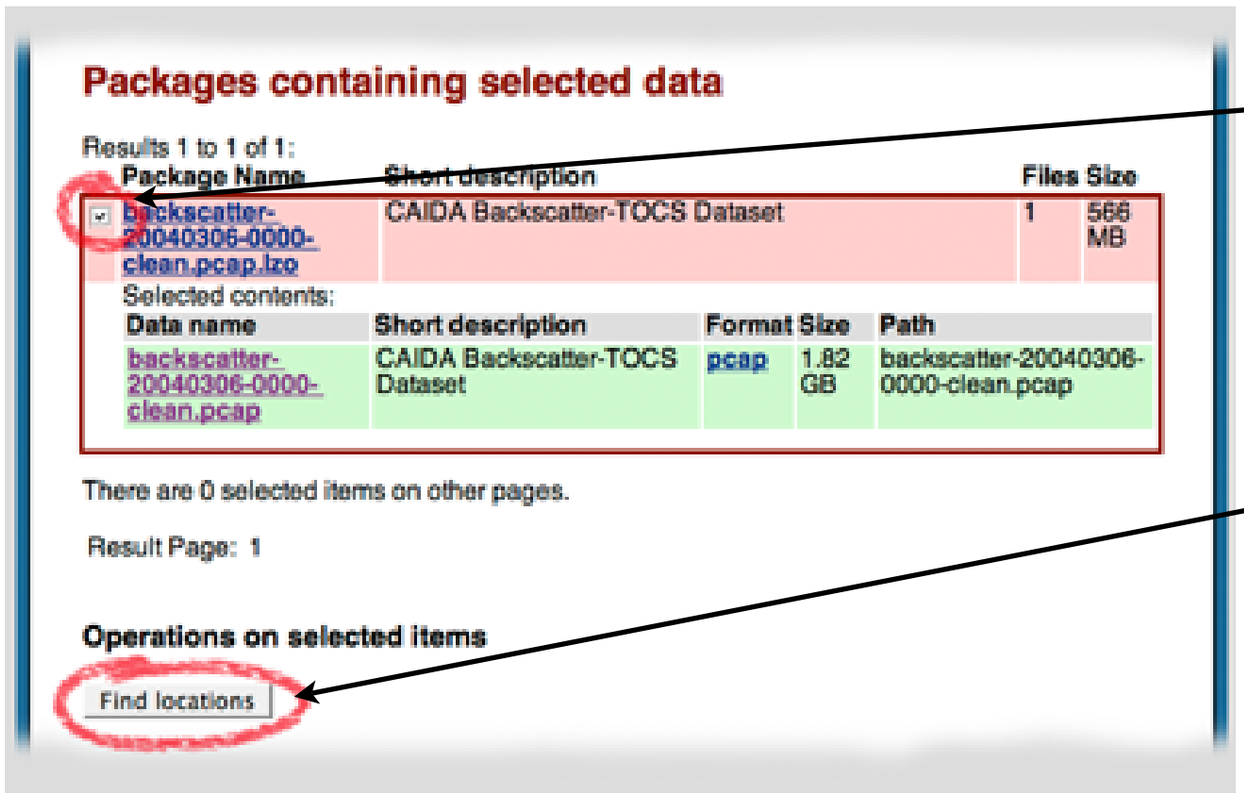
select desired data

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

http://imdc.datcat.org



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

http://imdc.datcat.org

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