Two days in The Life of The DNS Anycast Root Servers

Ziqian Liu
Beijing Jiaotong Univeristy

Bradley Huffaker, Marina Fomenkov
Nevil Brownlee, and kc claffy
CAIDA

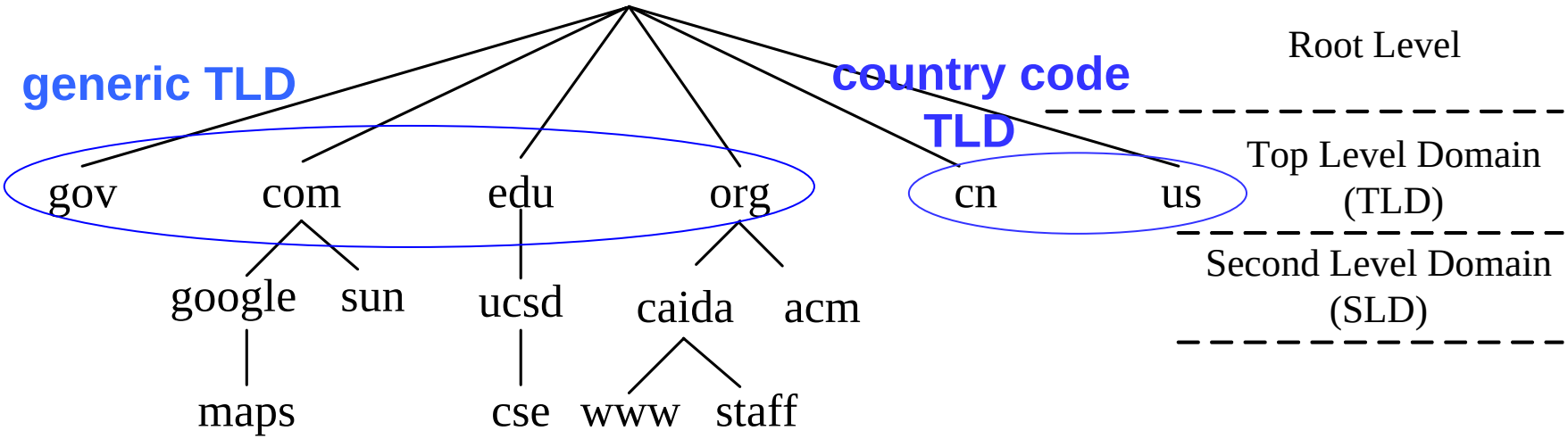
PAM2007

Outline

- **DNS root servers**
- **DNS anycast in root servers**
- **Data**
- **Traffic difference in anycast instances**
- **Anycast coverage**
- **Conclusion**

DNS Root Servers

- **Tree-structured distributed database**



- **Root servers provide authoritative referrals to name servers for gTDL and ccTLD domains.**

- **Only 13 root servers world wide**

  `[A-M].root-servers.net`

DNS Anycast in Root Servers

- **What is anycast**
  - **Anycast group**
    - **A set of instances that are run by the same organization and use the same IP address – the *service address* – but are physically different nodes.**
    - **e.g., k.root-servers.net – RIPE – 17 instances – 194.0.14.129**
  - **For a DNS root servers, anycast provides a service where by clients send requests to the service address and the network delivers that request to at least one, preferably the closest, instance in the root server's anycast group.**

DNS Anycast in Root Servers (2)

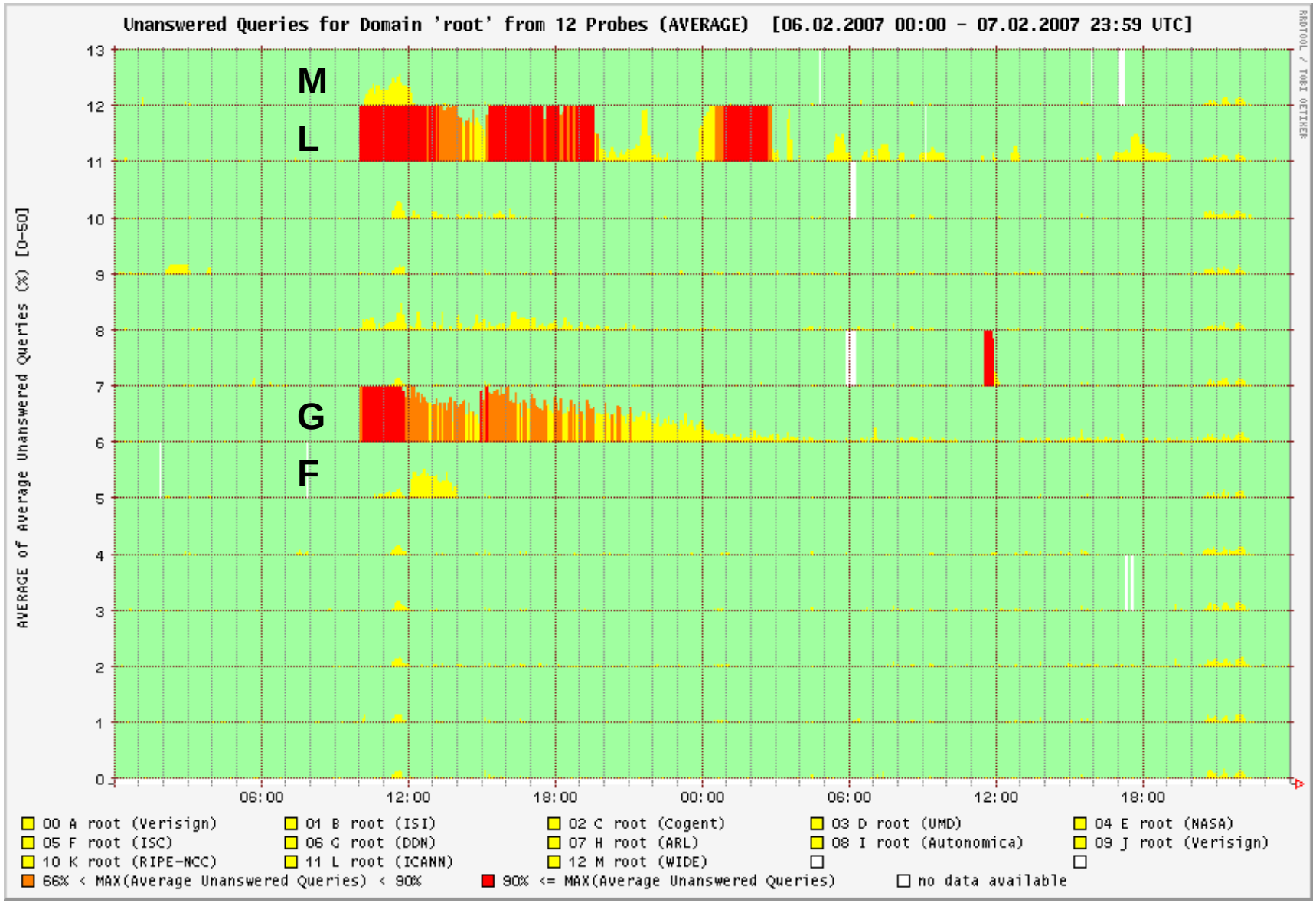
- **How to deploy**
  - **Every instance in the anycast group announces reachability for the same prefix – *service supernet* – that covers the service address by BGP.**
    - **e.g `k.root-servers.net` `193.0.14.0/24`**
    - **So, multiple AS paths are advertising the same prefix.**
  - **Different BGP routing policies:**
    - **Local instance**
      - **Limit the catchment area by using `no-export` community**
    - **Global instance**
      - **Globally visible**
      - **Use AS-prepending to decrease the likelihood of their selection over a local instance**

DNS Anycast in Root Servers (3)

- **C, F, I, J, K and M roots (6/13) have deployed anycast.**
  - **Over 120 root instances all togther (www.root-servers.org)**

- **What's the benefit?**
  - **Allow the system to grow beyond the static 13 servers while avoiding a change to the existing protocol**
  - **Bring DNS service closer to the clients**
  - **Provide relatively reliable and stable service compared to a non-distributed structure.**
    - **Separate the server/network failure**
    - **Mitigate the impact of malicious traffic**
      - **Oct 2002 DDoS attack against 13 root servers**
      - **Feb 2007 DDoS attack against root servers and gTLD servers**

http://dnsmon.ripe.net/



Data

- **ISC/ OARC (DNS Operations and Analysis Research Center) /CAIDA have been conducting measurement at the DNS root servers**

- **Three anycast root servers:**
  - **C-root: 4 of 4 instances**
  - **F-root: 33 of 37 instances (40 up-to-date)**
  - **K-root: 16 of 17 instances**

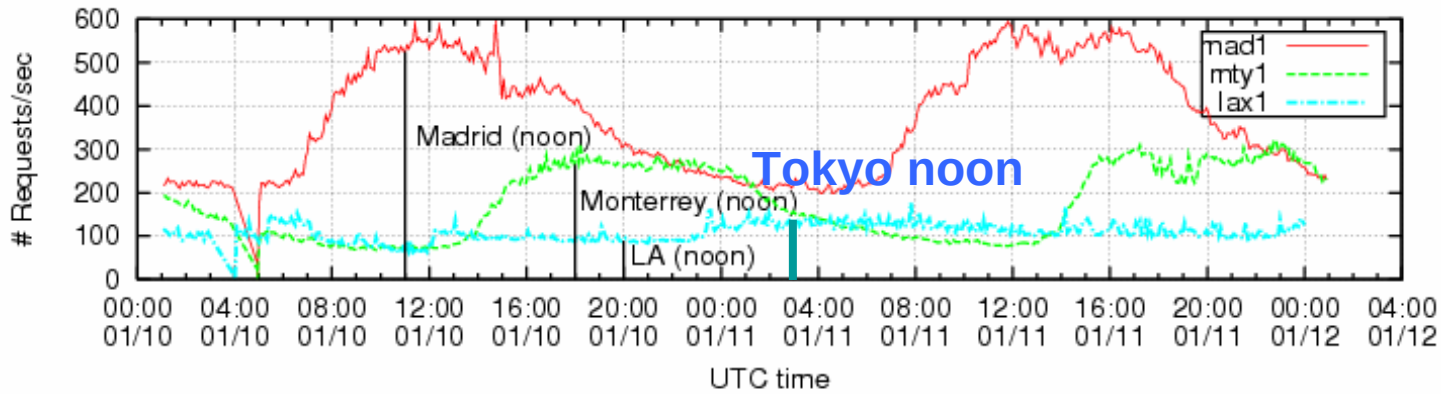
- **Time**
  - **Tue Jan 10~Wed Jan 11, 2006, UTC,**
  - **47.2 hour long (~2 days)**

- **Data format**
  - **Full record tcpdump traces**

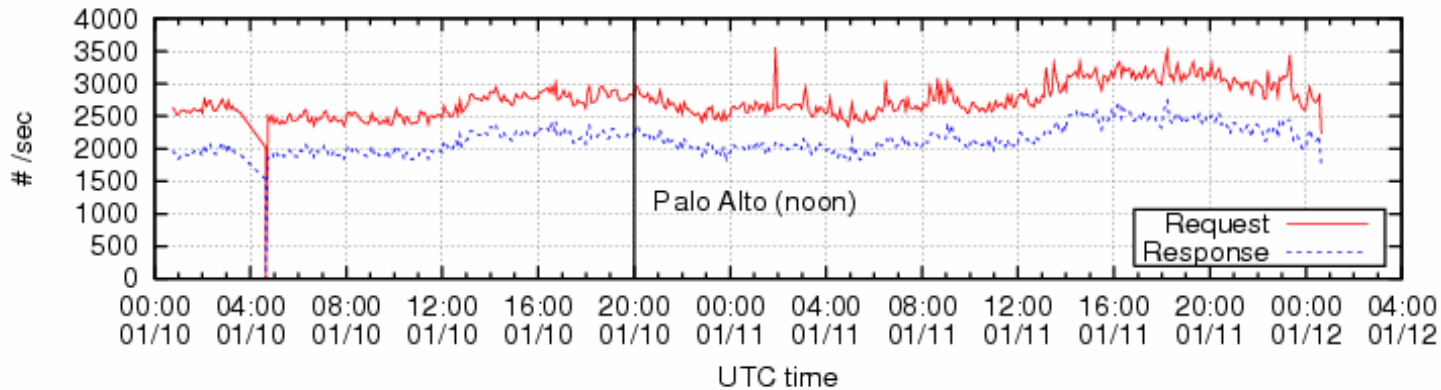
- **Our Focus: IPv4 UDP DNS requests**

- **Available at http://imdc.datcat.org**

Traffic difference – **Diurnal pattern**



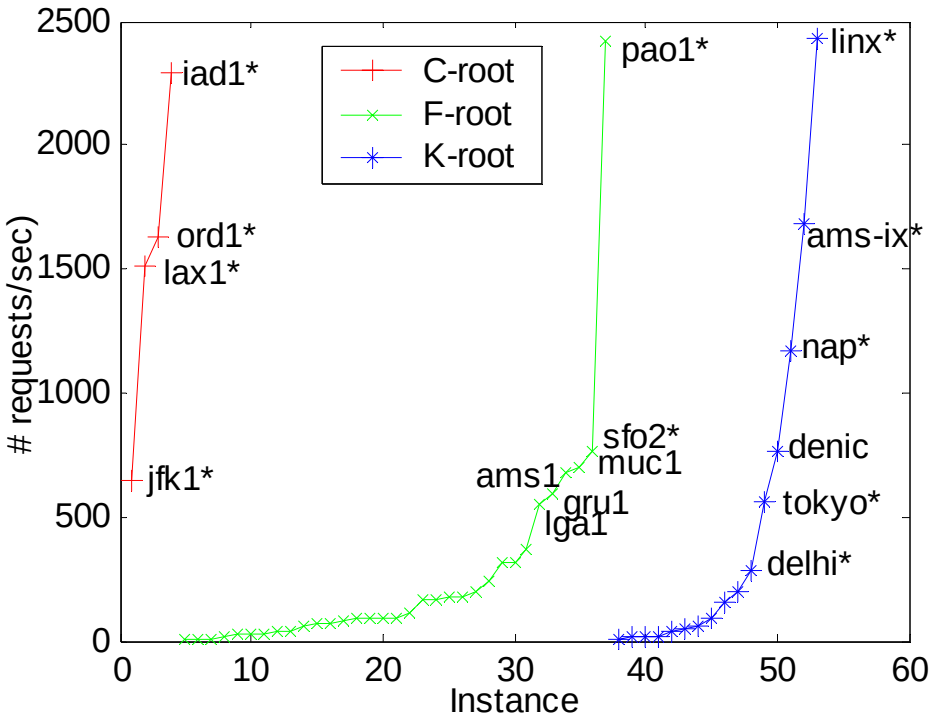
(a)The DNS request rates of three F-root local instances **mad1** (Madrid, Spain), **mty1** (Monterrey, Mexico), and **lax1** (Los Angeles, US).



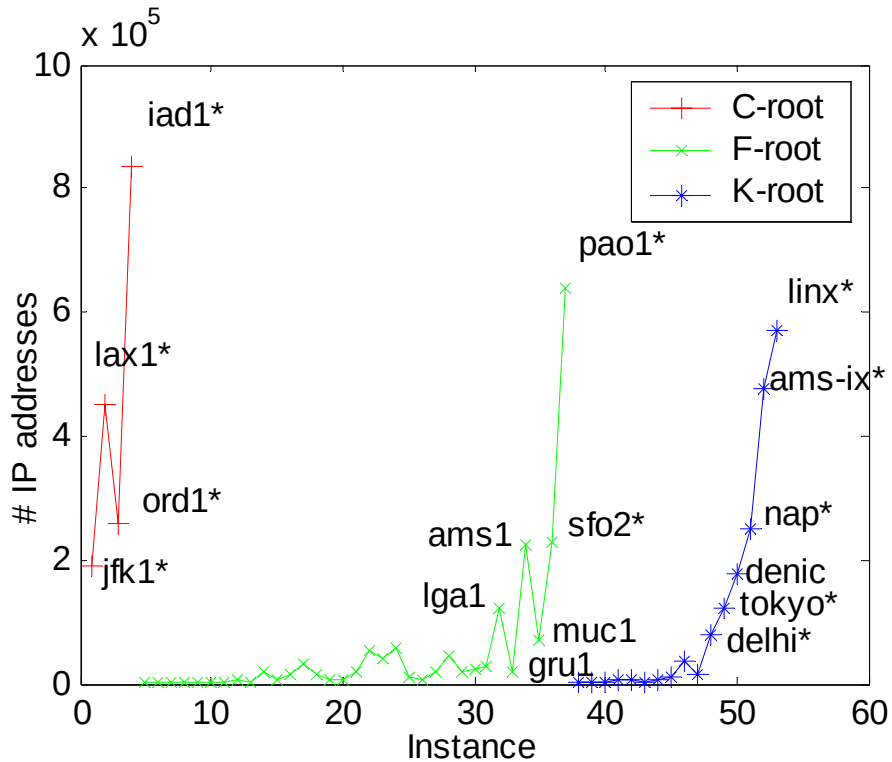
(b)The DNS request and response rates of the F-root global instance **pao1** (Palo Alto, US).

back

Traffic difference – Traffic load



(a) Average request rate

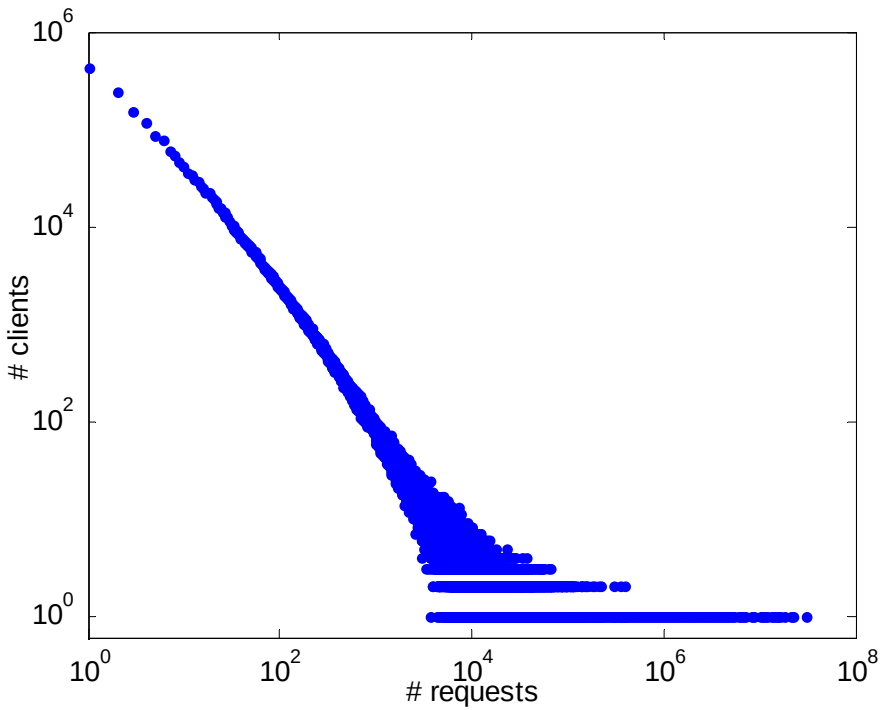


(b) Number of clients

Note

- **Both plots have the same x-axis intance order – instances within each group are arranged in an increasing request rate order**
- **Global instances are marked with ***

General statistics – **Clients vs. Requests**



During the 2 days:

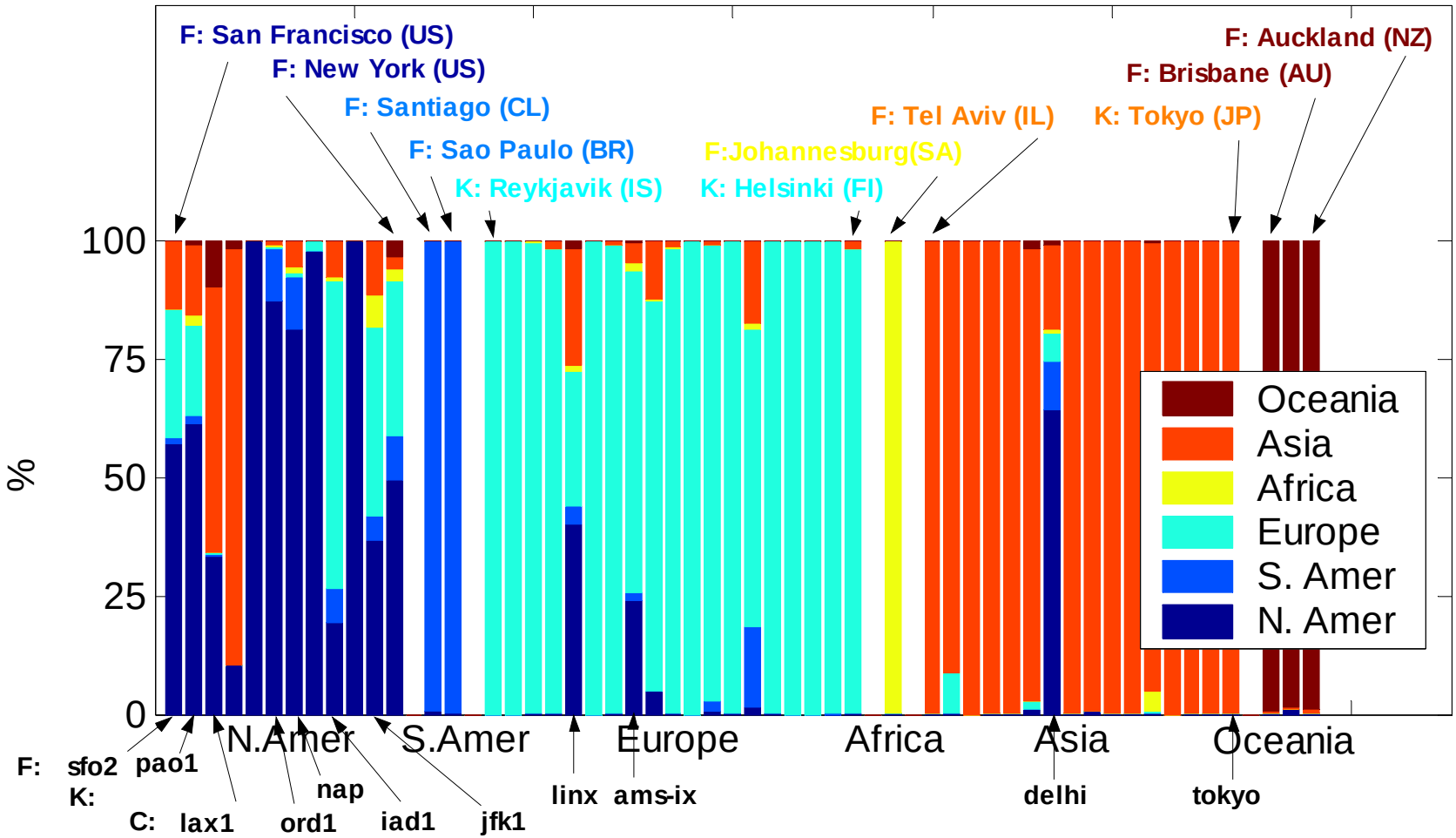
- **80% of the 2.5M clients sent <100 reqs to the three roots**
- **15 clients sent > 10M reqs**
- **Top client sent > 30M reqs i.e. ~ 174reqs/sec !**

Anycast coverage

- **Geographic**
  - **Client location: map client IP address to Geo info by using NetAcuity database**
  - **Instance location: coordinates of the closest airport**
  - **Distance: great circle distance**
- **Topological:**
  - **Route Views BGP table on Jan 10, 2006 for ASN and prefix**

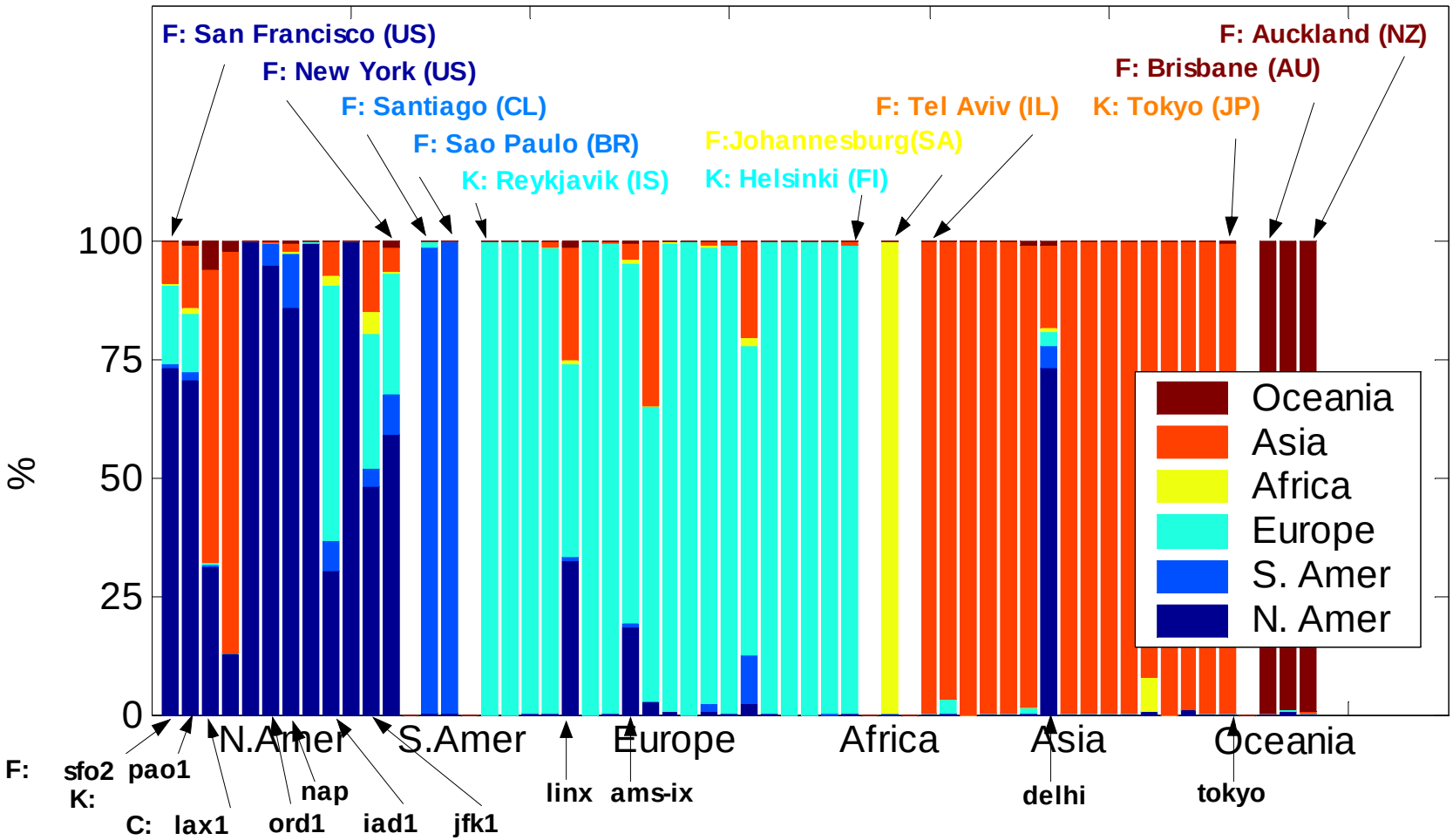
Anycast coverage – geographic distr.

- **Client Continental distribution**



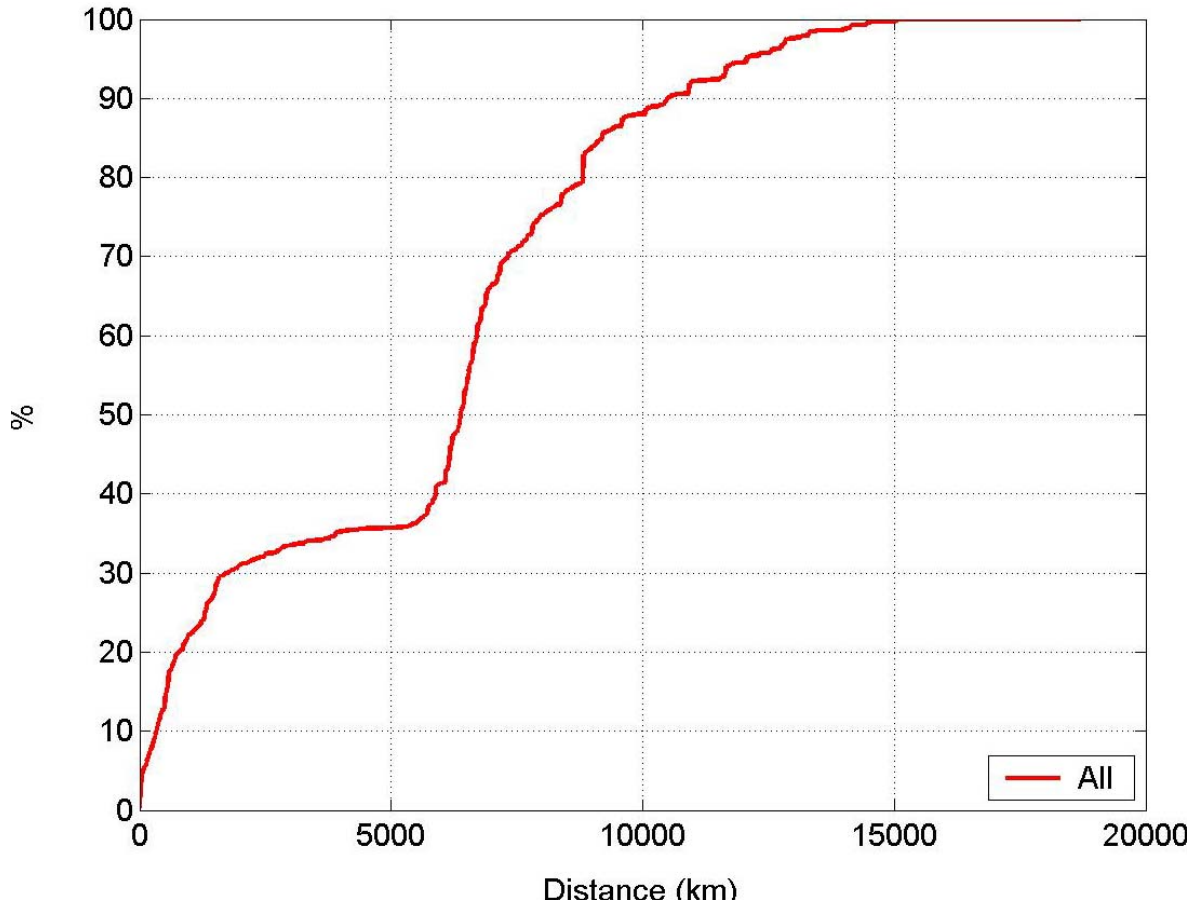
Anycast coverage – geographic distr.

- **DNS request continental distribution**



Anycast coverage – geographic distr.(2)

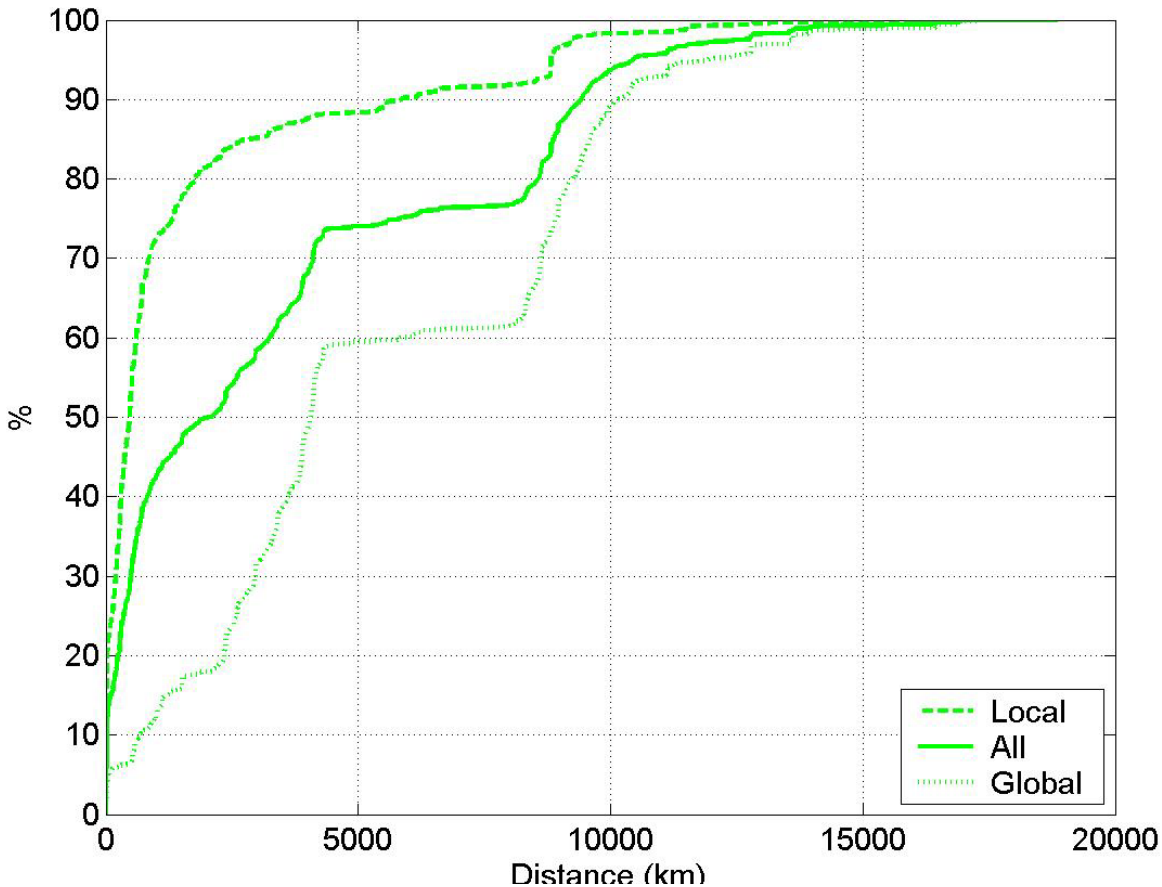
- **Distance distribution (instance ⟷ client)**



C-root

Anycast coverage – geographic distr.(3)

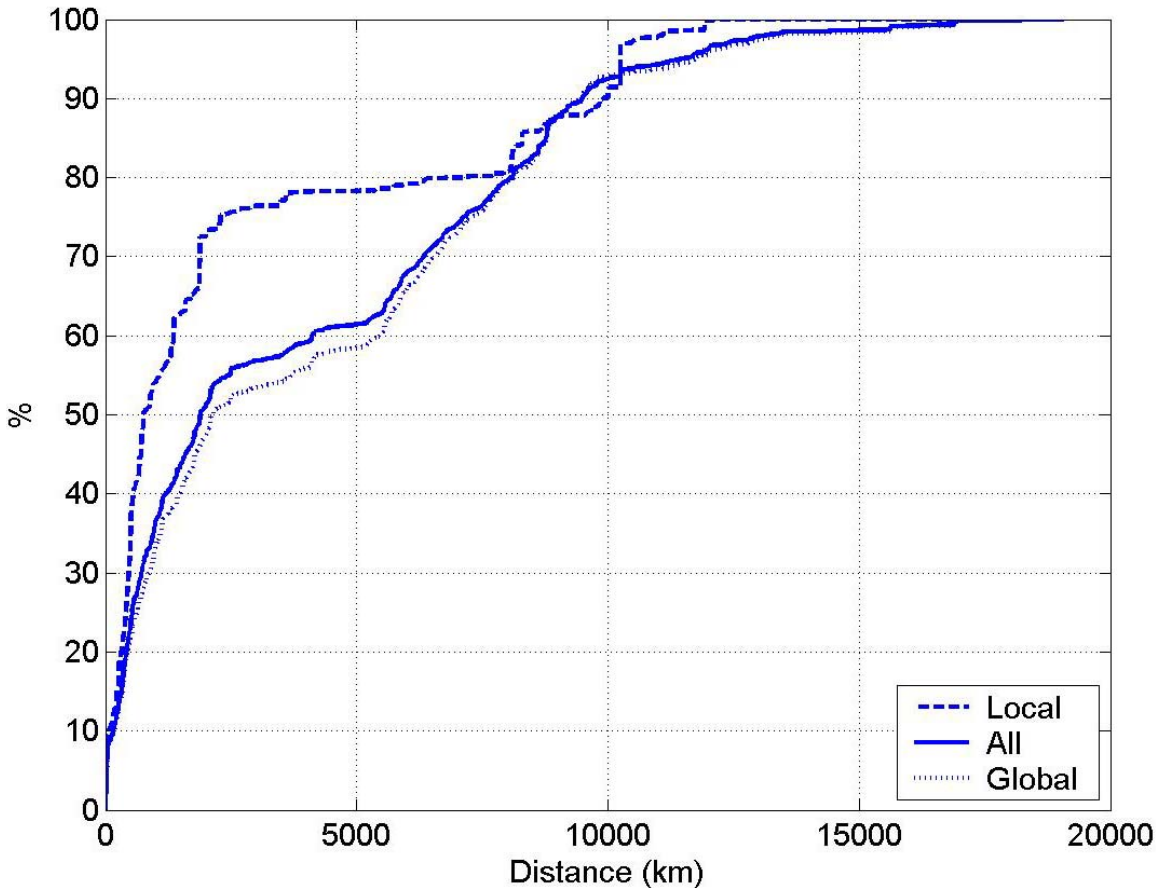
- **Distance distribution (instance ⟷ client)**



F-root

Anycast coverage – geographic distr.(4)

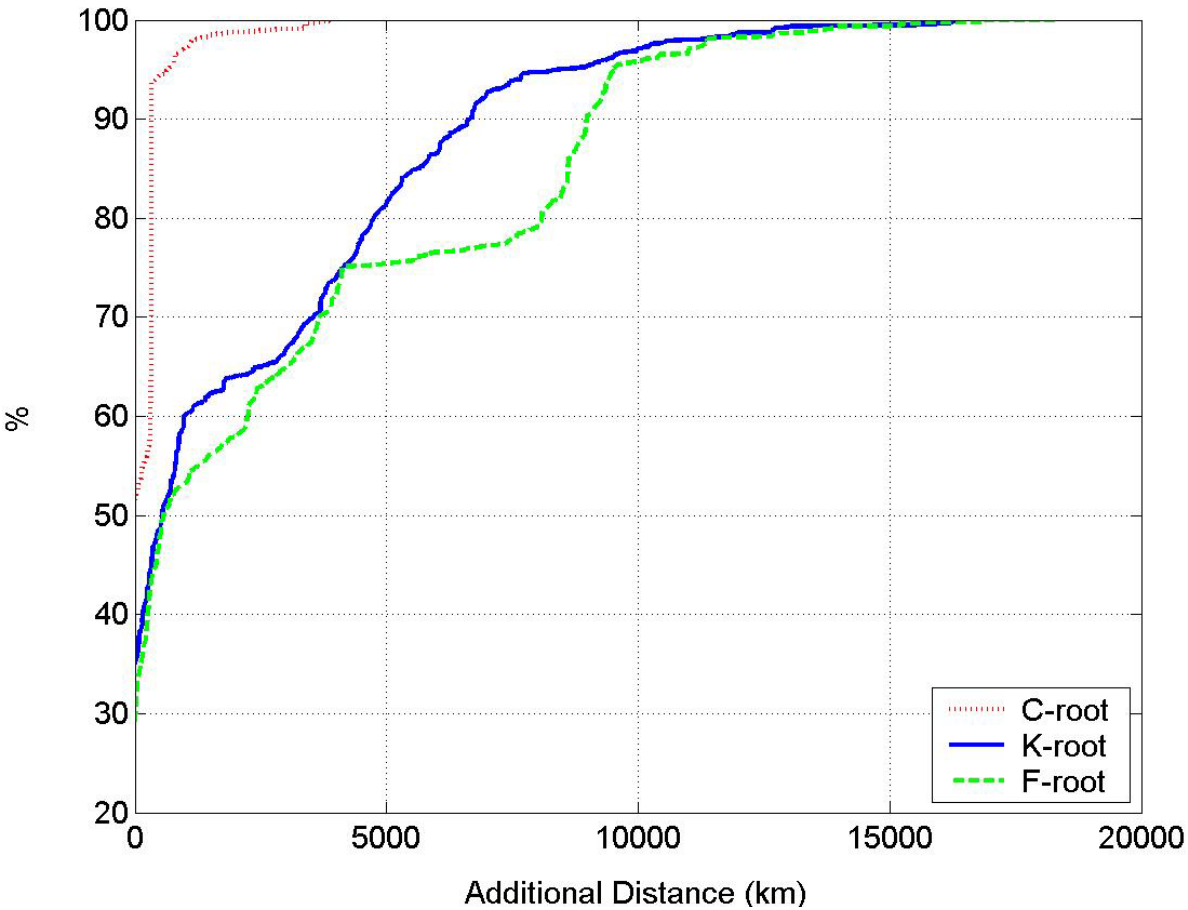
- **Distance distribution (instance ⟷ client)**



K-root

Anycast coverage – geographic distr.(5)

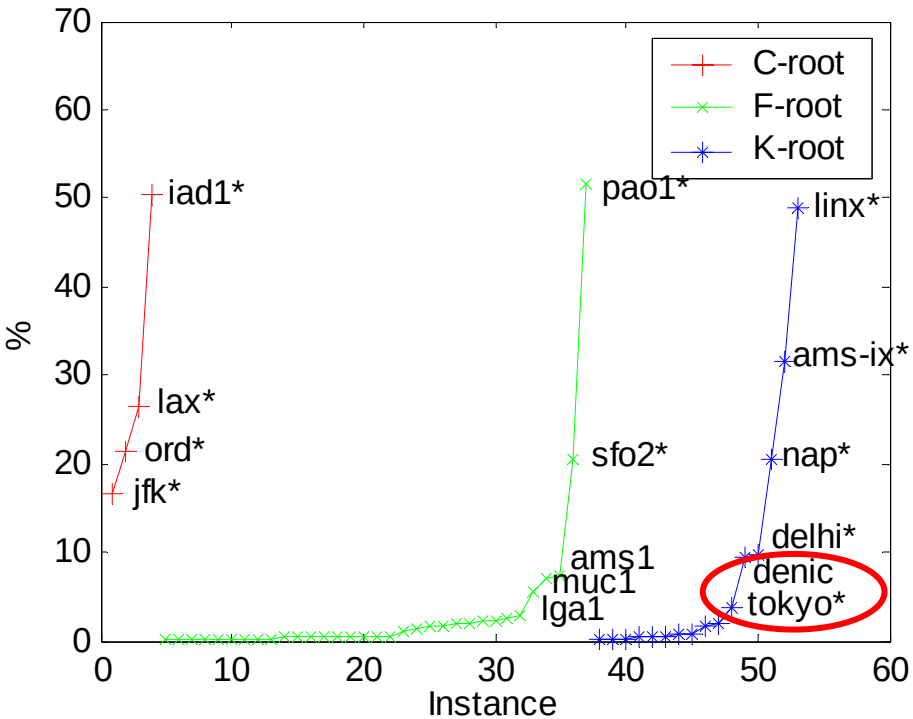
- **Additional distance** = **distance from the client to the instance it requests – distance from the client to the closest instance**



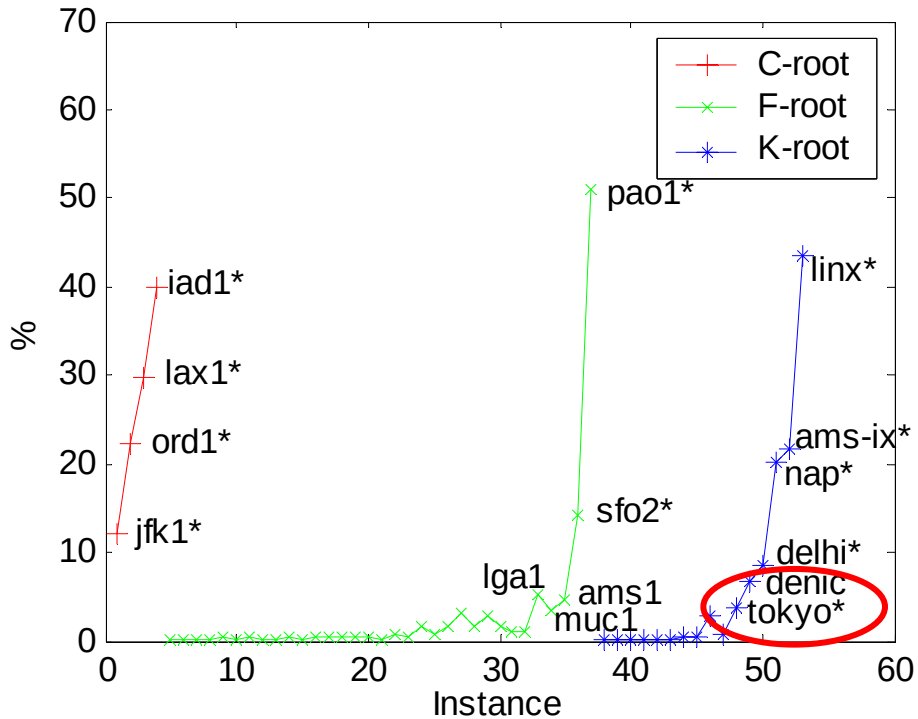
Anycast coverage – topological coverage

Topological scope: observed 19,237 ASes, RouteViews 21,883 ASes (~88%)

% = # seen by instance / # seen by all



AS level

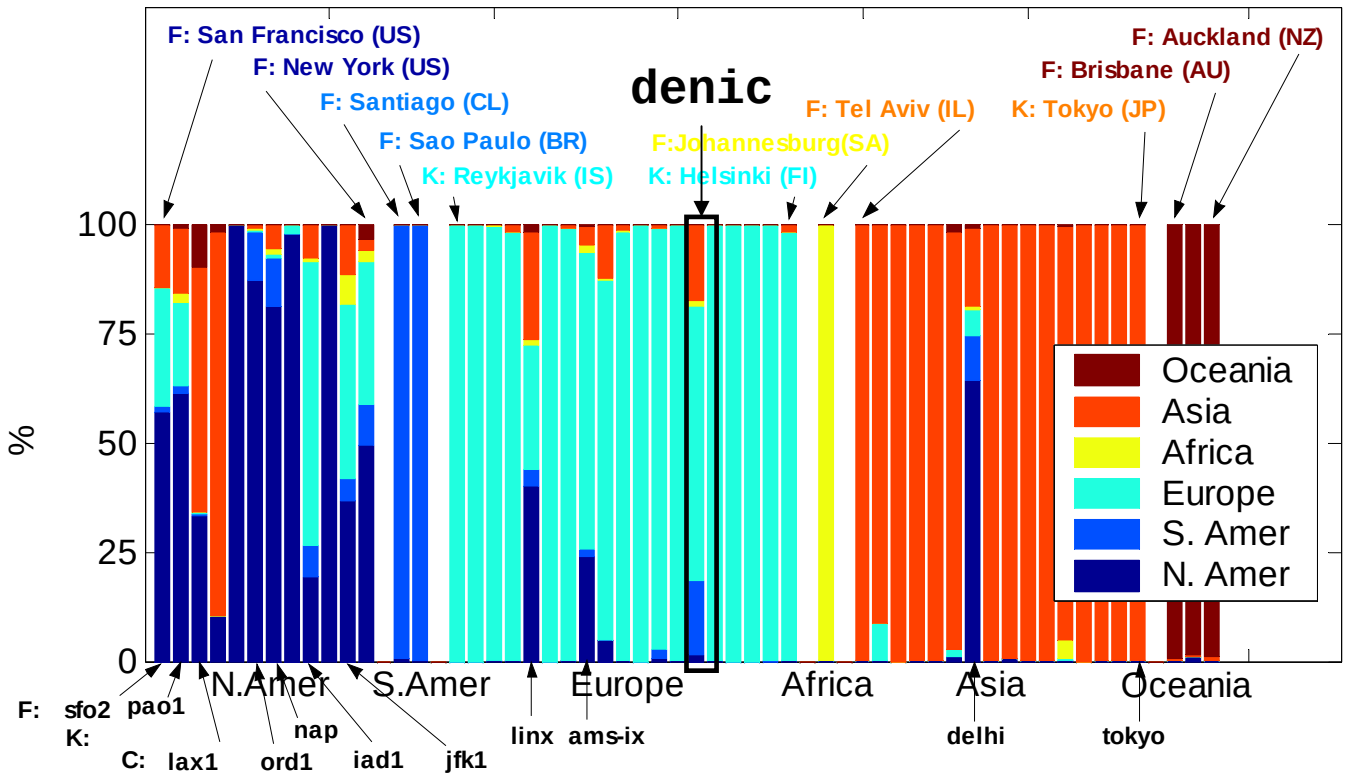


Prefix level

Both plots have the same x-axis intance order – instances within each group are arranged in an increasing AS coverage percentage order

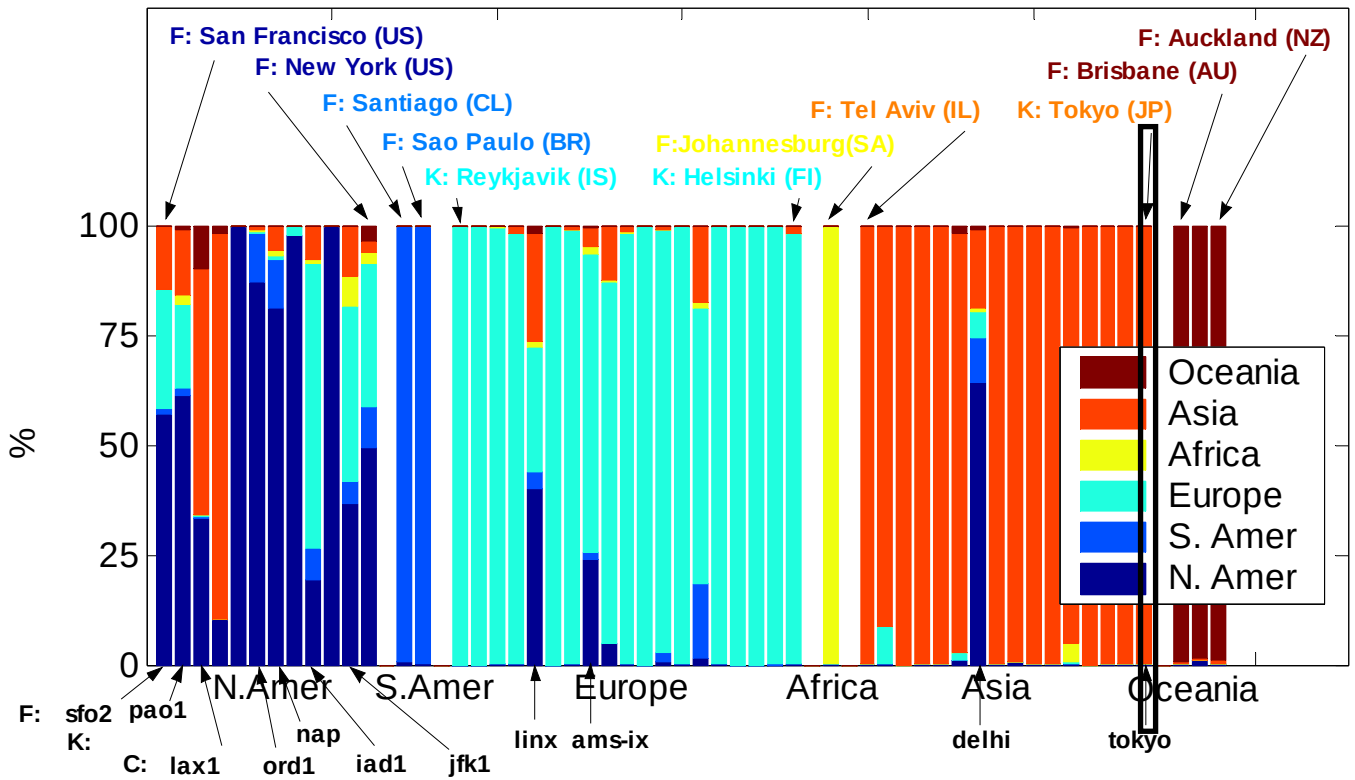
Anycast coverage – topological coverage (2)

- **denic – K-root local instance in Frankfurt, Germany**
  - **AS paths observed from RouteViews:**
    193.0.14.0/24 “3292 8763 25152 i”
    “4513 8763 25152 i”
    “12956 8763 25152 i”
  - **AS12956 belongs to Telefonica which has a large-scale coverage**



Anycast coverage – topological coverage (3)

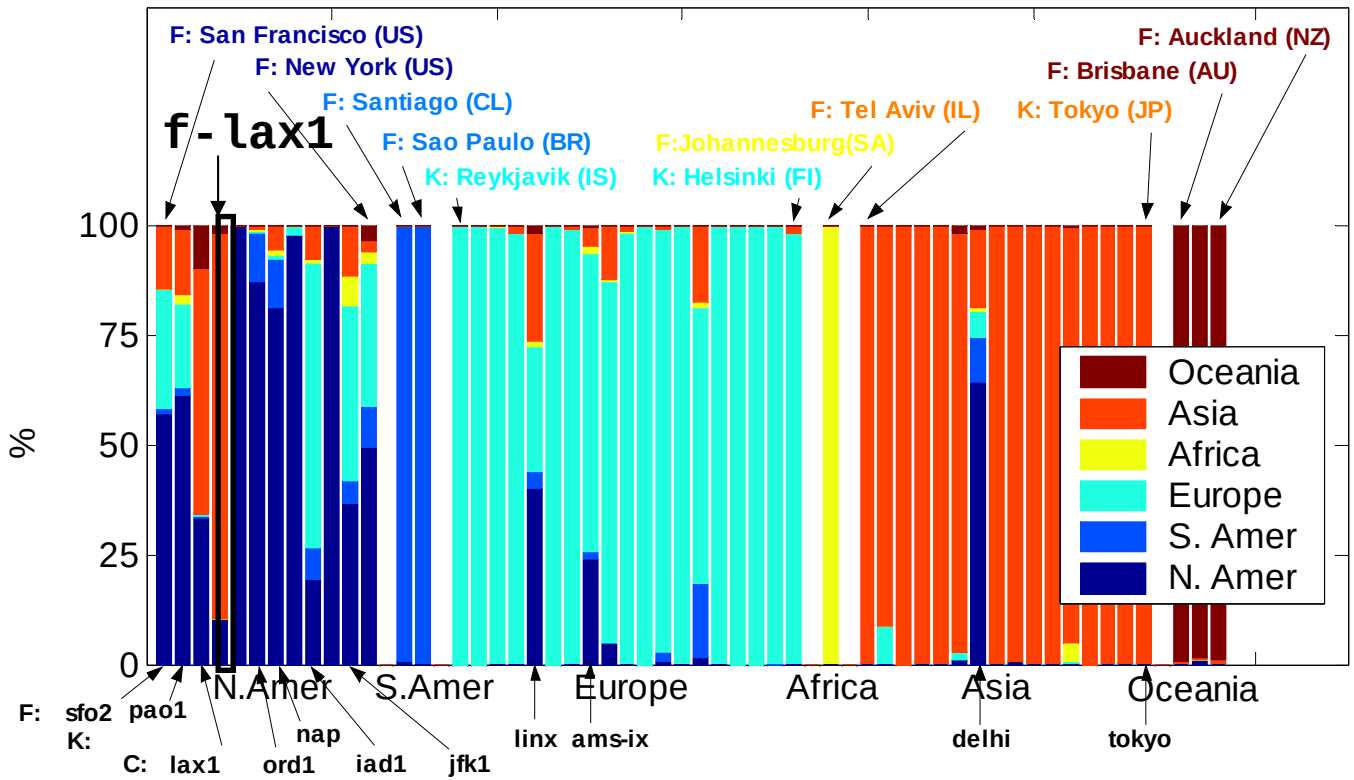
- **tokyo – K-root global instance in Tokyo, Japan**
  - **AS paths observed from RouteViews:**
    193.0.14.0/24 “4713 25152 25152 25152 25152 i”
  - **The longest among all five K-root global instances !**



Anycast coverage – topological coverage (4)

- **lax1 – F-root local instance in LA, U.S.**
  - **AS paths observed from RouteViews:**
    192.5.5.0/24 “7660 2516 27318 3557 i”
    “7500 2497 27318 3557 i”
    “2497 27318 3557 i”

    Both AS7660 and AS2516 are in Japan!



go

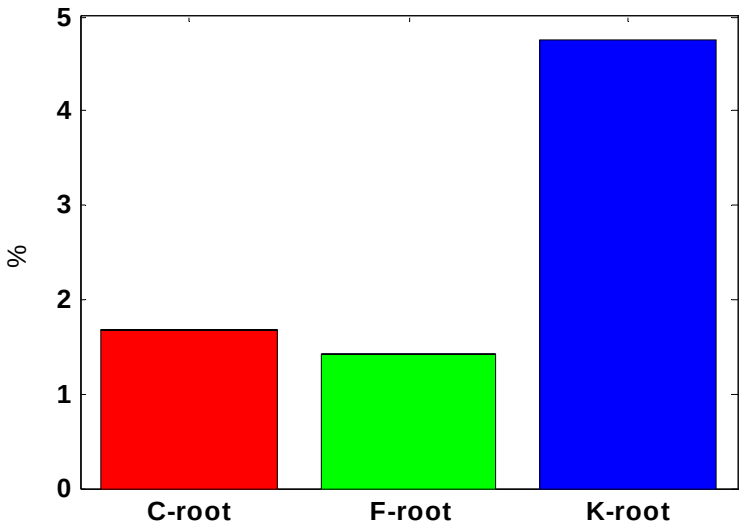
Anycast coverage – instance affinity

- **Anycast improves stability by shortening AS paths**
- **Anycast increases chance of inconsistency among instances and of clients' transparent shifting to different instances.**
- **Given DNS traffic is dominated by UDP, route flapping is unimportant**
- **But if DNS uses stateful transaction(TCP, fragmented UDP)…[Barber,NANOG32]**
- **Recent studies [Lorenzo] [Boothe,APNIC19] [Karrenberg,NANOG34]suggests the impact of routing switches on the query performance is rather minimal**

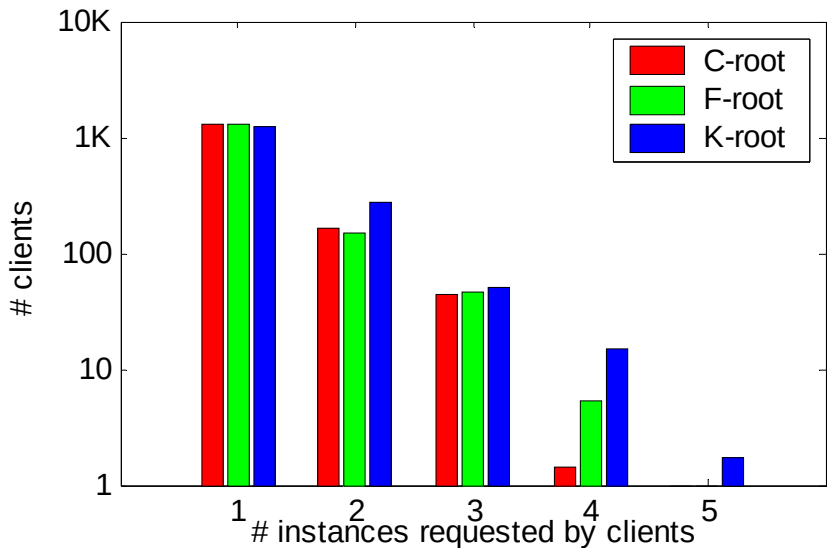
Anycast coverage – instance affinity (2)

- **2 F-root global instances together saw 99.8% of total clients who switched F-root instances**
- **5 K-root global instances together saw 86% of total clients who switched K-root instances**
- **Focusing on the clients who switched the most instances:**
  - **2 C-root clients are from Brazil and Bolivia**
  - **3 K-root clients are from Uruguay**
  - **27 F-root clients all from UK and they never requested `lcy1` – the F-root local instance in London, instead they switch between `ams1`, `lga1`, `pao1` and `sfo2`**

All from South America, no C or K root instance there!



Pct. of clients switching instances



Conclusion

- **Current method for limiting the catchment areas of local instances appears to be generally successful (though exceptions do exist due to peculiar routing configuration)**

- **A significant number of clients would benefit if routing configurations of their local DNS root instances were optimized to route them to the closet instances**

- **Instance selection by BGP is highly stable**
  - **In 2-day period, <2% of both C and F-root clients and < 5% of K-root clients experienced an instance change.**

Questions?