DNS Measurements

Hua Huang
CS8803 NTM

Outline

- "DNS measurements at a root server"
  Nevil Brownlee, Kimberly Claffy, and Evi Nemeth,
  Proceedings of the IEEE GlobeCom, San Antonio, TX, Nov. 2001
- "DNS Performance and the Effectiveness of Caching"
  J. Jung and E. Sit and H. Balakrishnan and R. Morris,
  IEEE/ACM Transactions on Networking, V10, n5, October, 2002
- "Diversity in DNS Performance Measures"
  R. Liston and S. Srinivasan and E. W. Zegura,
  Proceedings Internet Measurement Workshop (IMW), Nov., 2002
- "On the problem of optimization of DNS root servers' placement"
  Tony Lee, Bradley Huffaker, Marina Fomenkov, kc claffy,
  Passive and Active Measurement Workshop (PAM), La Jolla, CA, 2003

Outline

- "DNS measurements at a root server"

  Nevil Brownlee, Kimberly Claffy, and Evi Nemeth,

  Proceedings of the IEEE GlobeCom, San Antonio, TX, Nov. 2001

Measurement Methodology

- Passive measurement
  - Observe DNS traffic flowing to and from F root name server (*F.root-servers.net*, *located at PAIX*)
  - Using *tcpdump* to capture the entire DNS packets
- Data Collection

Size	Queries	Distinct Q's (%)	Date/Time	Collection Length
3.6 GB	10.3 M	2.7 M (26.2%)	Jan 7, 11 am	1 hour
5.9 GB	18.0 M	4.8 M (26.7%)	Jan 9, 3 pm	2 hours
10.4 GB	29.1 M	4.5 M (15.5%)	Jan 8, 1 pm	1 hour
338 MB	1 M	380 K (37.9%)	Jan 10, hourly	2M pkts (4 minus)
690 MB	2 M	622 K (31.2%)	Jan 12,17–19,24	4M pkts (8 mins)

- Access to a full set of error logs
  - Denied attempts to dynamically update the root server
  - Dropped queries that were received with source port 0

Query Rates

- Use *netstat* command to measure raw query rate
- Data
  - 01/06/2001 – 01/16/2001, 01/25/2001 – 01/31/2001
- Work week query peak load – 5000/sec
- 93% of the queries were responded immediately
- 7% unanswered quires
  - Quires from private address space – no route back
  - Malformed queries
    - 256 queries specified in the header, with only 1 query actually
      - Big endian, little endian byte order problem in the nameserver code on some NT4/Win95/Win98 machines
      - Jan. 7, 2001 (1 hr trace): 78,000 queries from 1400 distinct nameservers with this bug

Errors Identified

- Repeated queries
  - Not understand *referral* or *SERVFAIL* responses

    Average: 154 times per second
  - 01/20/2001

    a single host repeated over 2 million times in an hour (SERVFAIL)
- Private address space (RFC 1918)
  - 2-3% of queries arriving at F root have source IP in RFC 1918 space
  - 7% queries ask for hostnames in RFC 1918
  - 7% queries from an RFC 1918 address ask about such an address
- Invalid top level domains (TLDs)
  - 01/07/2001 trace (1 hr)
    - 16.5% of the servers asked only invalid queries
  - Spelling errors
  - Local nameserver add local domain to complete the name

    www.bcs.WSCOOPER.WSCOOPER.... Until 255 characters long

Errors Identified (cont'd)

- Bogus *A* queries
  - Over 14% of root server's query load is due to queries that violate DNS specification
    - 12%-18% queries with an IP address as a target
    - Causes identified
      - Win2K resolver library, *snow white* virus, *wininit* virus
      - OpenBSD resolver and some DSL modem boxes
- Source port zero
- Dynamic Updates
  - Requests trying to update root servers
    - When *Win2K* was 1st released, it flooded the root servers with requests to update the root zone

Other Anomalies

- Denial of service attacks
  - Use root as a reflector, flooding the attack target with answers to questions it did not ask
  - Scan the IP space but did not reverse IP address bytes when querying for an associated hostname

    199.170.0.2.1024 PTR 54.11.193.155.in-addr.arpa.
- Microsoft’s DNS problem
  - Put all of their externally visible nameservers on the same subnet
  - 01/24/2001, router misconfiguration at Microsoft caused load on root for MS names to increase from 0% to 25%

Outline

- "DNS measurements at a root server"

  Nevil Brownlee, Kimberly Claffy, and Evi Nemeth,

  Proceedings of the IEEE GlobeCom, San Antonio, TX, Nov. 2001

- "DNS Performance and the Effectiveness of Caching"

  J. Jung and E. Sit and H. Balakrishnan and R. Morris,

  IEEE/ACM Transactions on Networking, V10, n5, October, 2002

Data Collection

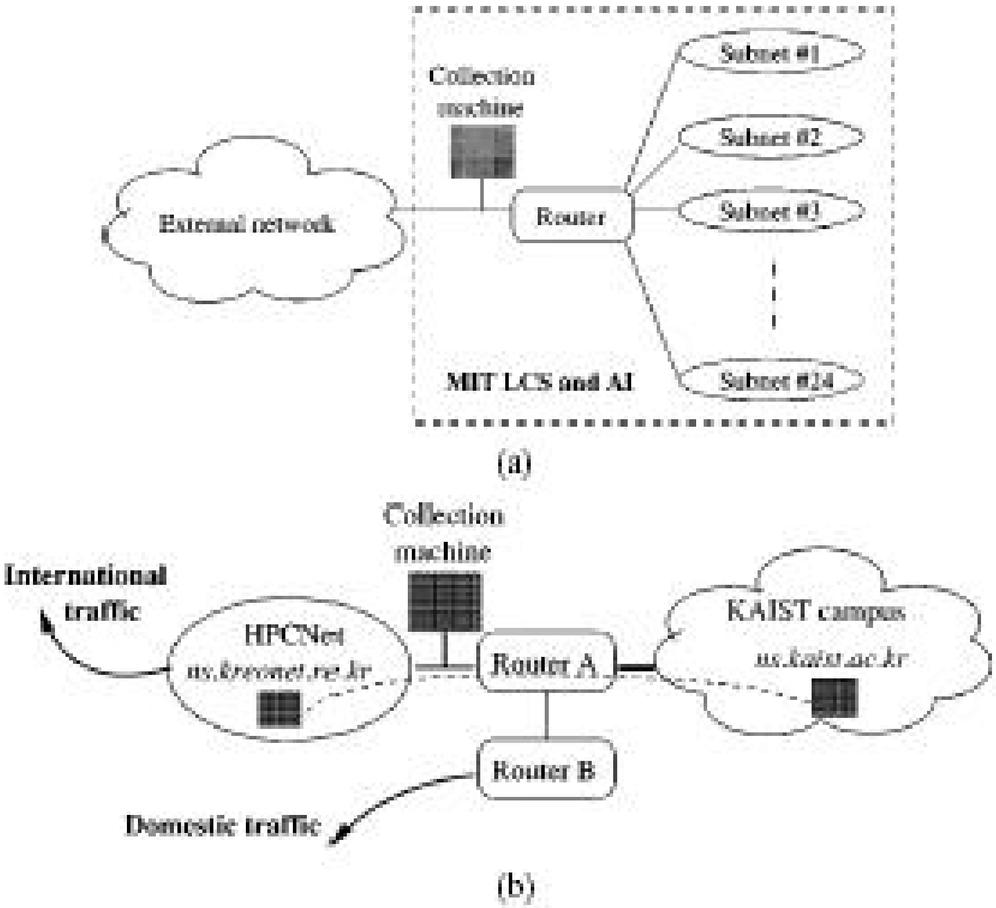


Fig. 2. Schematic topology of the traced networks

(a) MIT LCS: There are 24 internal subnetworks sharing the border router.

(b) KAIST: The collection machine is located at a point that captures all DNS traffic, but only international traffic of other types.

- Collection software
  - Derived from Minshall's *tcpdpriv* utility
- Both DNS traffic and its driving TCP connections
- 3 traces
  - mit-jan00
    2:00am 01/03/00-01/10/00
  - mit-dec00
    6:00pm 12/04/00-12/11/00
  - Kaist-may01
    5:00am 01/03/00-01/10/00

Terminology

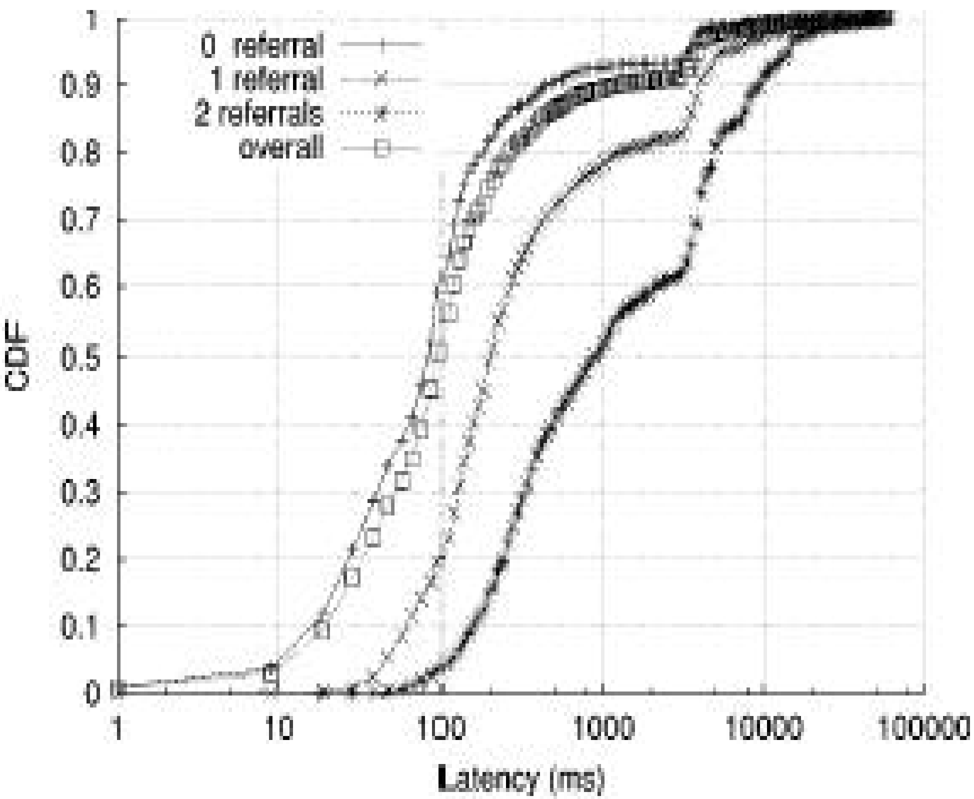
- *Lookup*
  - Entire process of translating a domain name for a client application
- *Query*
  - A DNS request packet sent to a DNS server
- *Response*
  - A packet sent by a DNS server in reply to a query packet
- *Answer*
  - A response from a DNS server that terminates the lookup, by returning a requested mapping or an error indication
- *Zero answer*
  - Is authoritative and indicates no error, but has no ANSWER, AUTHORITY, or ADDITIONAL records

Results

		mit-jan00	mit-dec00	kaist-may01
1	Date and place	00/01/03-10, MIT	00/12/04-11, MIT	01/05/18-24, KAIST
2	Total lookups	2,530,430	4,160,954	4,339,473
3	Unanswered	595,290 (23.5%)	946,308 (22.7%)	873,514 (20.1%)
4	Answered with success	1,627,772 (64.3%)	2,648,025 (63.6%)	1,579,852 (36.4%)
5	Answered with failure	281,855 (11.1%)	545,887 (13.1%)	1,834,942 (42.2%)
6	Zero answer	25,513 (1.0%)	20,734 (0.5%)	51,165 (1.2%)
7	Total iterative lookups	2,486,104	4,107,439	396,038
8	Answered	1,893,882	3,166,353	239,874
9	Total query packets	6,039,582	10,617,796	5,326,527
10	Distinct second level domains	58,638	84,490	78,322
11	Distinct fully-qualified names	263,984	302,032	219,144
12	Distinct internal query sources	221	265	405
13	Distinct external name servers	48,537	61,776	8,114
14	TCP connections	4,521,348	5,347,003	665,361
15	#TCP : #valid A answers, *sans* black-lists	4.62	3.53	–
16	Distinct TCP clients	992	1,233	5,754
17	Distinct TCP destinations	59,588	204,192	11,511

Table 1. Basic trace statistics. The percentages are with respect to total number of lookups in each trace.

Effect of Referrals on Latency

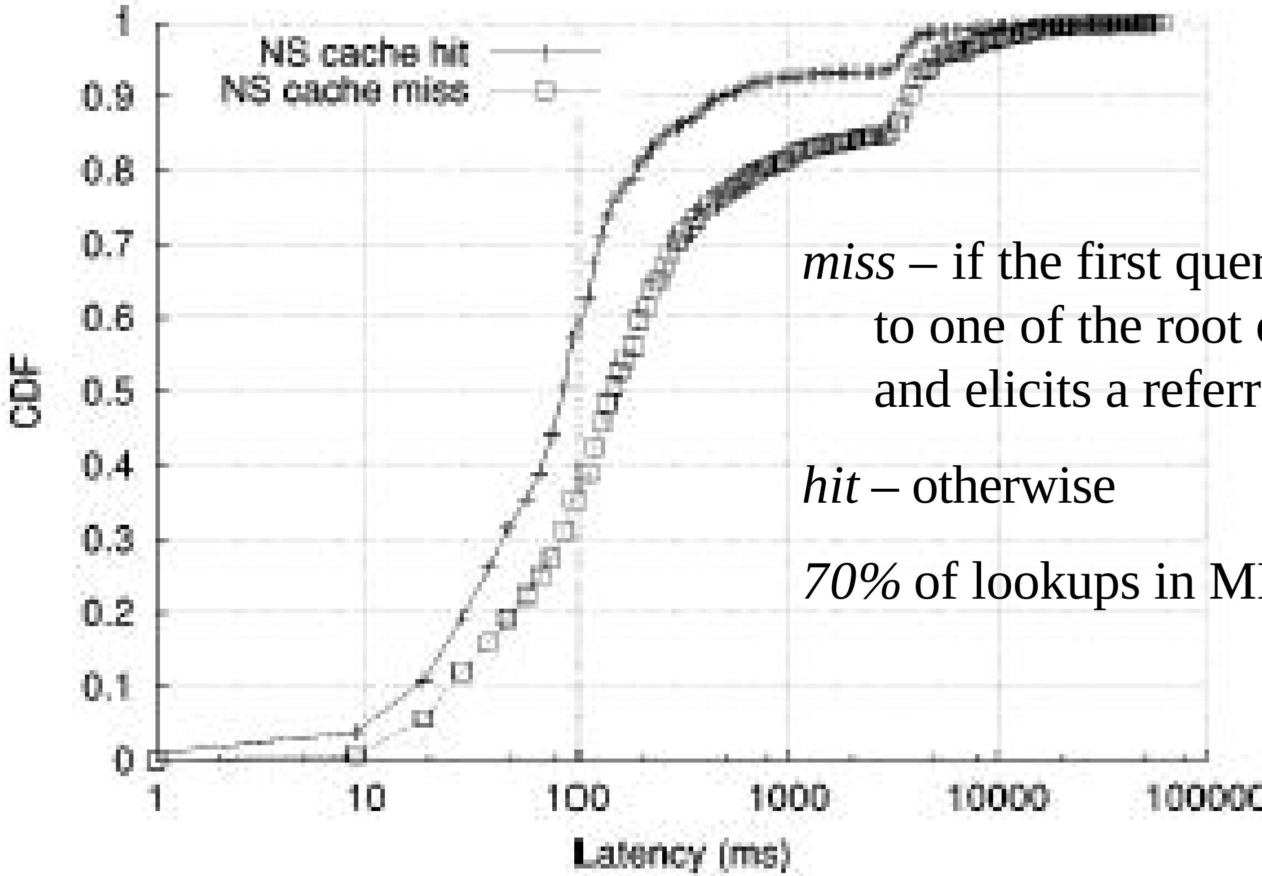


Latency distribution versus number of referrals for the mit-dec00 trace

Percentage of lookups involving various numbers of referrals

# of referrals	mit-jan00	mit-dec00	kaist-may01
0	74.62%	81.17%	86.09%
1	24.07%	17.86%	10.43%
2	1.16%	0.87%	2.10%
3	0.11%	0.07%	0.38%
= 4	0.04%	0.03%	1.00%

Effect of NS Records Caching on Latency



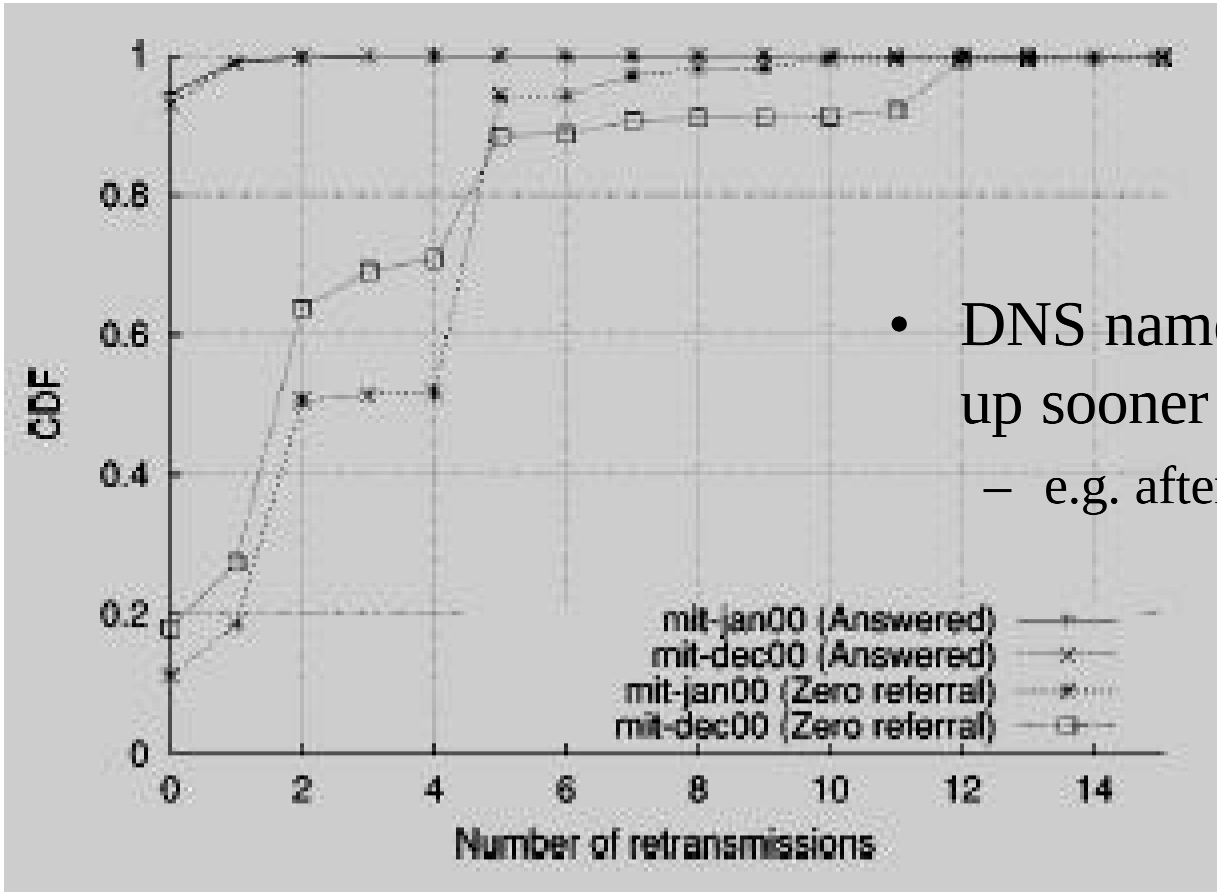
miss – if the first query packet is sent to one of the root or gTLD servers and elicits a referral

hit – otherwise

70% of lookups in MIT traces are hits

Distribution of latencies for lookups that do and do not involve querying root servers

Retransmission



- DNS name servers should give up sooner
  - e.g. after 2 or 3 retransmissions

Cumulative distribution of number of retransmissions for answered (topmost curves) and unanswered lookups

Negative Responses

Breakdown of Negative Responses by Cause as percentage of All Negative Responses

Cause	mit-jan00	mit-dec00
Non-existent name	82,459 (42%)	150,066 (32%)
No reverse map for PTR	79,725 (41%)	249,236 (54%)
No RBL (or similar) entry	11,552 (6%)	36,955 (7%)
Loopback	7,368 (4%)	11,310 (2%)
Other one-word names	4,785 (3%)	9,718 (2%)
Invalid characters in query	1,549 (1%)	5,590 (1%)

- Negative caching is not working as well as it could be
- Servers should not forward queries for unqualified names when resolving queries for the Internet class

Interaction with Root Servers

	mit-jan00	mit-dec00
Root Lookups	406,321 (16%)	270,413 (6.4%)
Root Errors	59,862 (2.3%)	73,697 (1.7%)
gTLD Lookups	41,854 (1.6%)	353,295 (8.4%)
gTLD Errors	2,676 (0.1%)	16,341 (0.3%)

Table 7. The total number of lookups that contacted root and gTLD servers, and the total number of failure answers received. The percentages are of the total number of lookups in the trace.

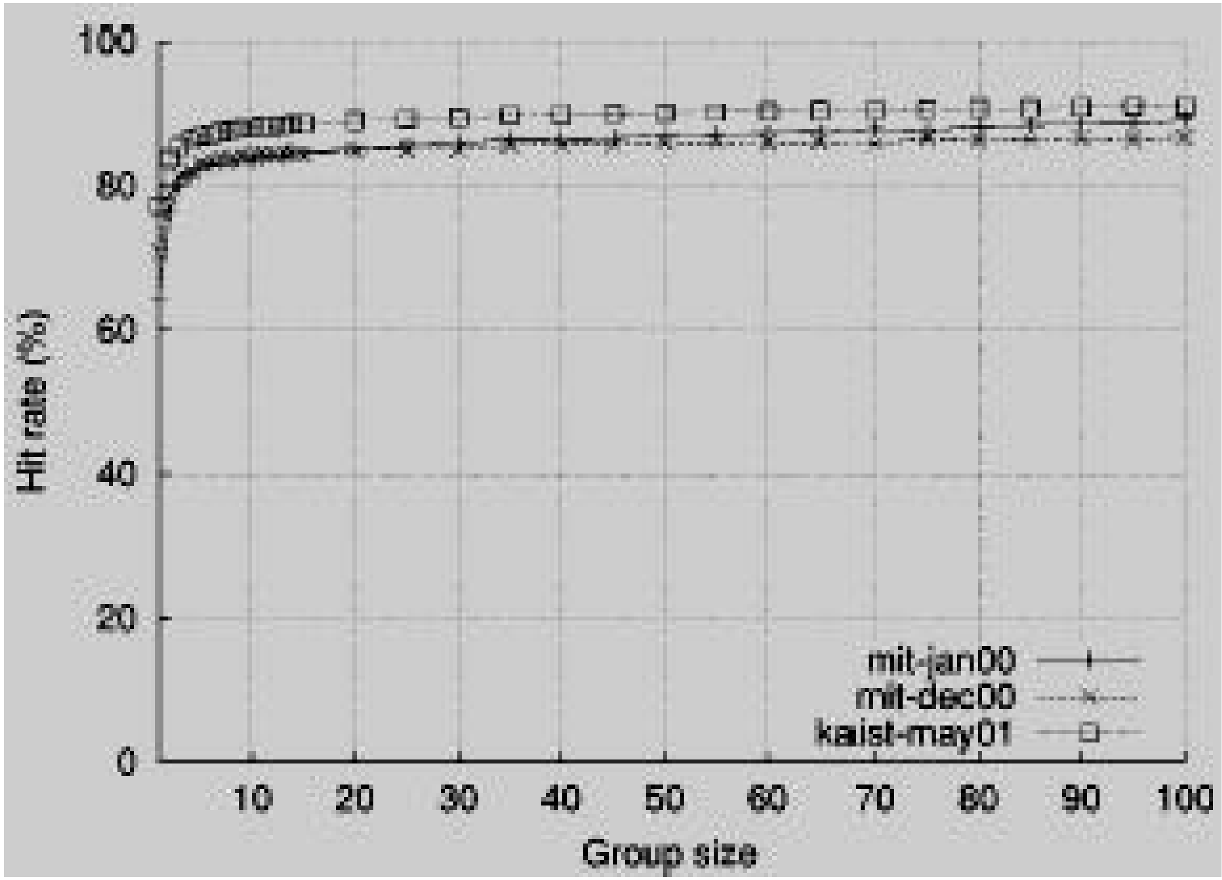
- 15% – 27% of lookups sent to root name servers resulted in negative responses
  - mistyped names, bare host names (*e.g.*, *loopback*), etc
  - Many of these are automatically generated by incorrectly implemented or configured resolvers

Effectiveness of Caching

- Trace-driven Simulation Algorithm
  - 2 databases: Name database TTL database
  - Simulation run
    1. Randomly divide TCP clients into groups of size ***s***. Each group has a simulated cache indexed by the domain name
    2. For each new TCP connection, find which group the client belongs to, check the cache entry, ***hit*** if the entry is there, ***miss*** otherwise

- Two issues
  - Usefulness to share DNS caches among many client machines
    - The extent to which different clients look up the same names
  - Impact of choice of TTL on caching effectiveness

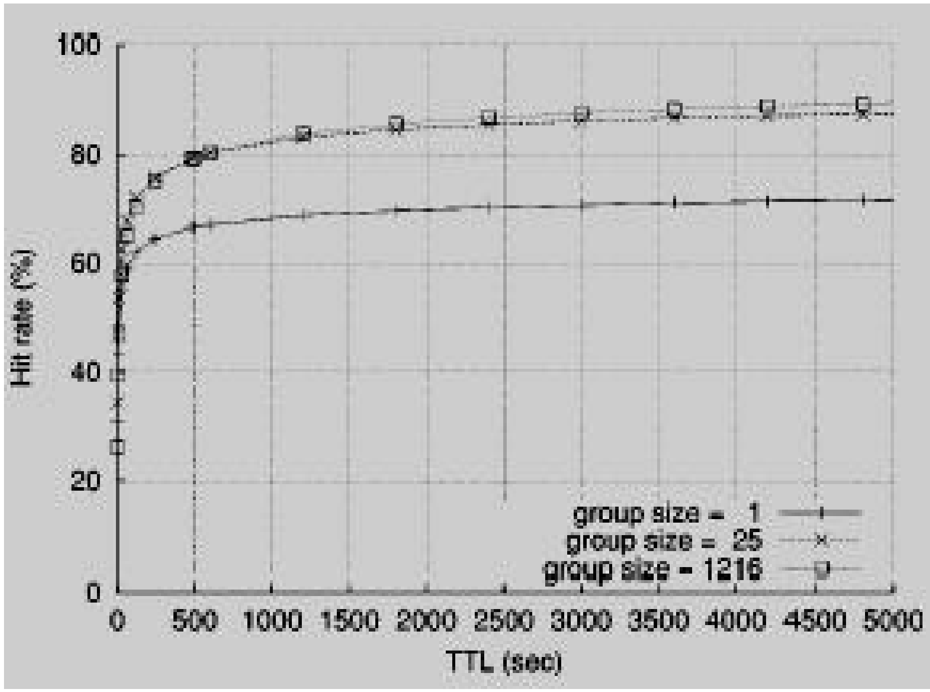
Effect of Sharing on Hit Rate



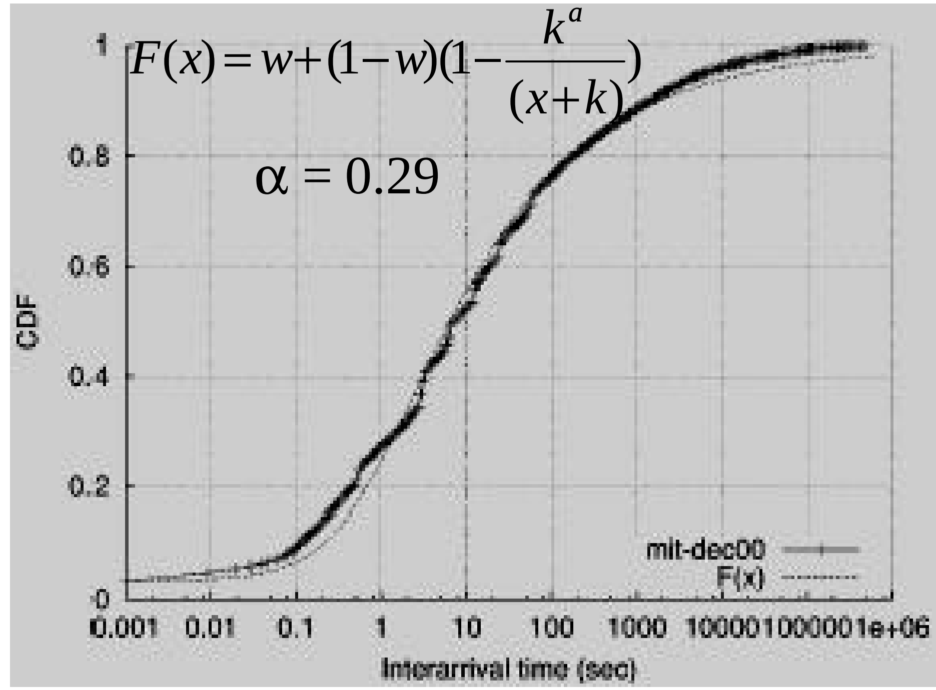
- Most of the benefits of sharing are obtained with as few as 10 or 20 clients per cache

Impact of TTL on Hit Rate

Impact of TTL on Hit Rate – mit-dec00



TCP Connection Interarrivals – mit-dec00



- Effect of TTL on the hit rate is noticeable only for
  TTL < 1000 sec
- Single clients look up same server multiple times in quick succession

- Pareto interarrival dist w/ point mass at $t = 0$
- $\alpha<1 \rightarrow$ infinite mean $\rightarrow$ limited additional benefits from longer TTL

Outline

- "DNS measurements at a root server"

  Nevil Brownlee, Kimberly Claffy, and Evi Nemeth,

  Proceedings of the IEEE GlobeCom, San Antonio, TX, Nov. 2001

- "DNS Performance and the Effectiveness of Caching"

  J. Jung and E. Sit and H. Balakrishnan and R. Morris,

  IEEE/ACM Transactions on Networking, V10, n5, October, 2002

- "Diversity in DNS Performance Measures"

  R. Liston and S. Srinivasan and E. W. Zegura,

  Proceedings Internet Measurement Workshop (IMW), Nov., 2002

Diversity in DNS Performance Measures

- Goal
  - Investigate the degree to which metrics for wide-area DNS performance differ across locations in the Internet
  - Non-cached domain names only
    - Users experience longest lookup times for non-cached names
- Metrics investigated
  - Completion and success rates of lookups
  - Mean response time for completed lookups
  - Root and gTLD servers favored by the sites
  - Observed fraction of names that are aliases
  - Distribution of TTLs across names
- Metrics expected to be invariant across locations

  Fraction of aliases, TTL distribution,

  Fraction of names that are successfully resolved

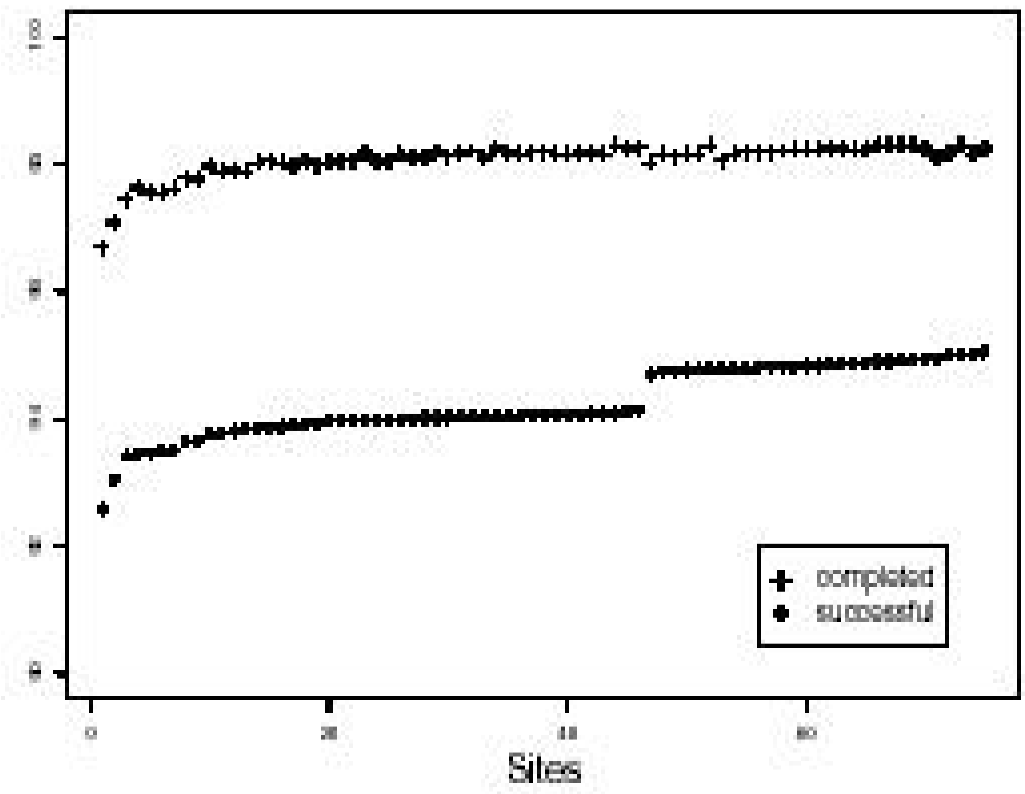
Measurement Methodology

- Measurement locations
  - 75 different Internet locations in 21 countries and territories
  - Various connection technologies

    DSL, PPP, cable modem, gigabit Ethernet, etc
  - Jan., 2002 and late March/early April, 2002
- Domain name sample
  - Obtained by crawling the web w/ *Larbin* crawler
  - 14,983 names w/ unique second-level domains to ensure measurements for non-cached names only
- Tool: *named* name server
  - log each event during name resolution w/ time stamp
  - 4-6 hrs of continuous operation to complete on each site
  - Bandwidth consumption

    Incoming: 5Kbps outgoing: 700bps

Completion and Success Rates



- Successful– return an answer with no error
- Complete – return an answer

- x-axis – sites ordered by the no. of successful lookups
- y-axis – % of completed and successful lookups

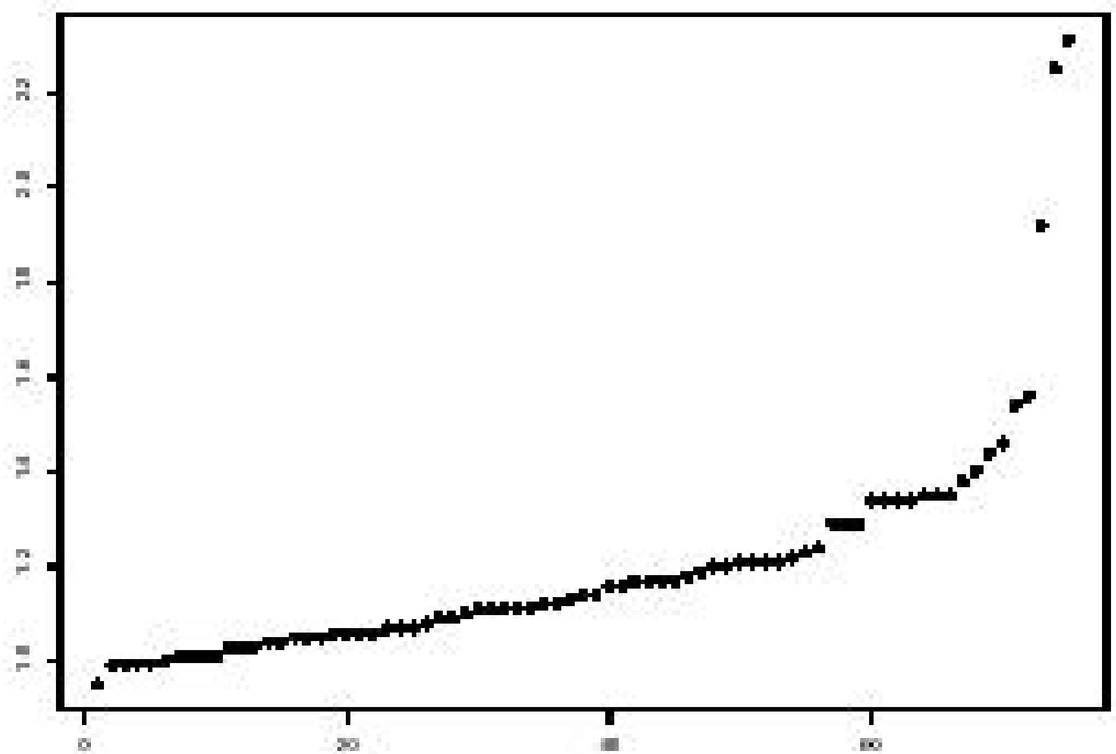
- Successful lookups – 2 clusters

  Jan. (higher rate), March/April
- 2 sites w/ slightly lower rates

  higher no. of retries for 3-16 mins

- Time-sensitive
- Possibly location-sensitive

Mean Response Time



x-axis – sites ordered by MRTc

y-axis – mean response time for completed lookups (sec)

- Large disparity in overall performance among each sites

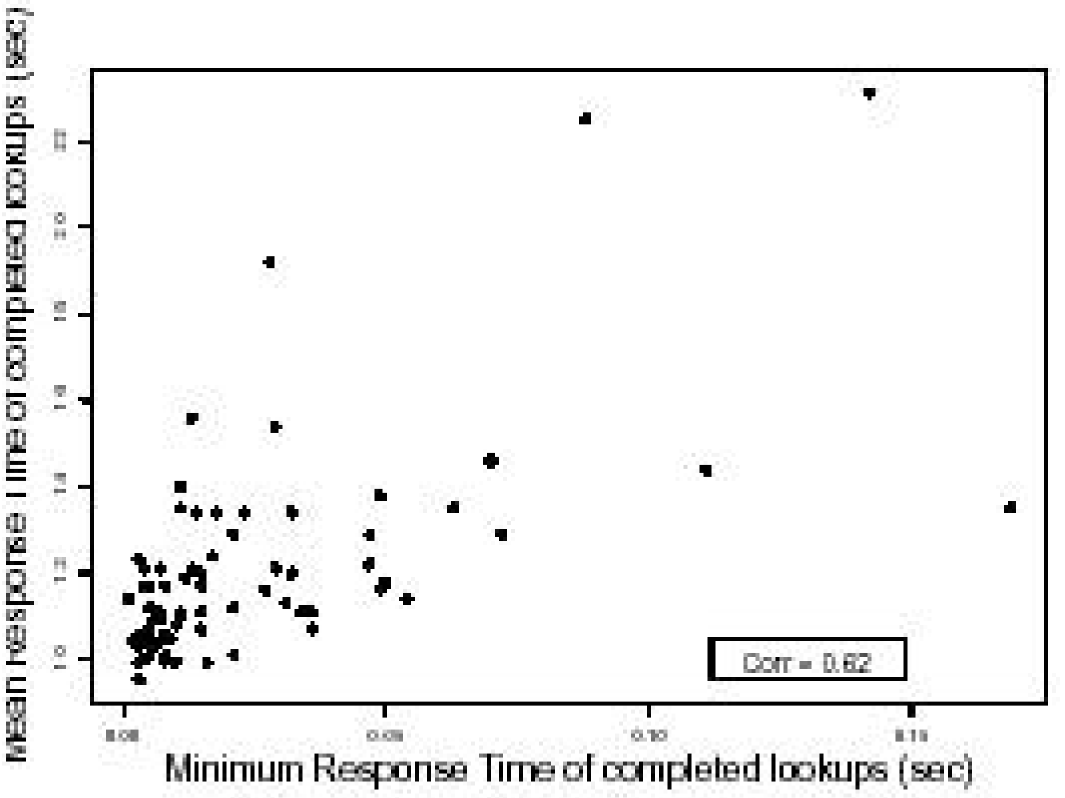
  max_MRTc / min_min_MRTc = 2.4

- Possible factors

  Connectivity, Loss rate, Perceived performance of root and gTLD servers

  Location in the network relative to other name servers

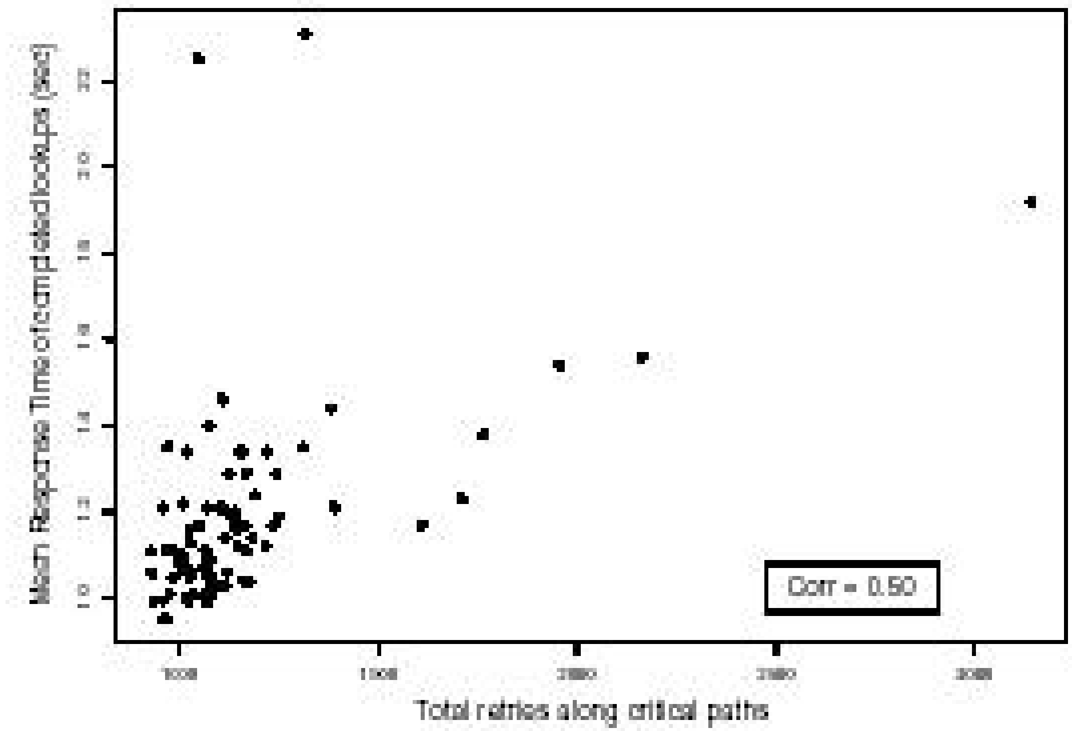
Connectivity



- Lower MINc → higher bandwidth connection and/or close proximity to the Internet
- Correlation ρ = 0.62

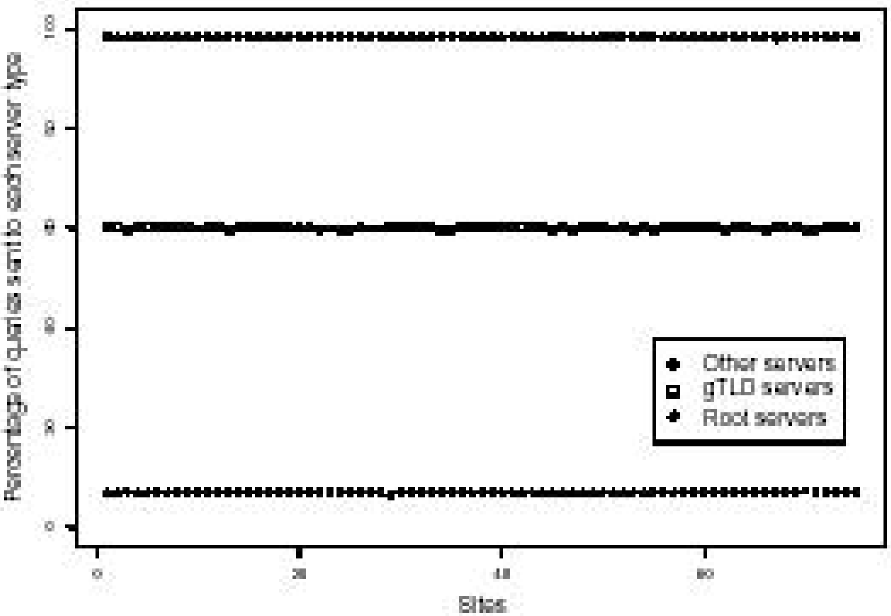
  Connectivity does not sufficiently account for the higher MRTc

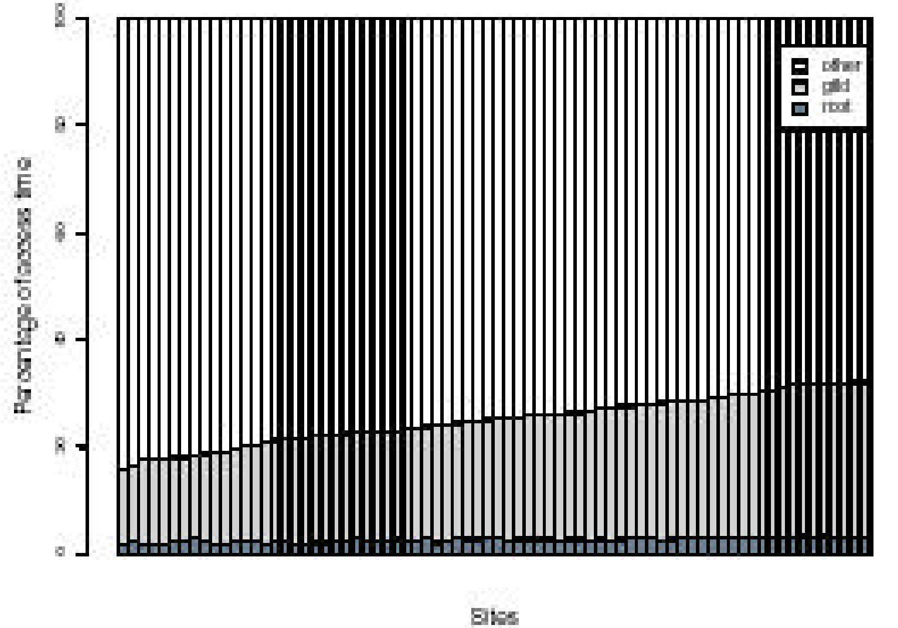
Loss Rate



- Correlation ρ = 0.50
- Assume that retries are a good measure of loss rate, loss rate is not a major factor affecting lookup time
- Loss rate varies dramatically across sites

Root/gTLD Server Performance





- Percentage of lookups where each type of server was queried

    Root: 7.0%, gTLD: 60.0%

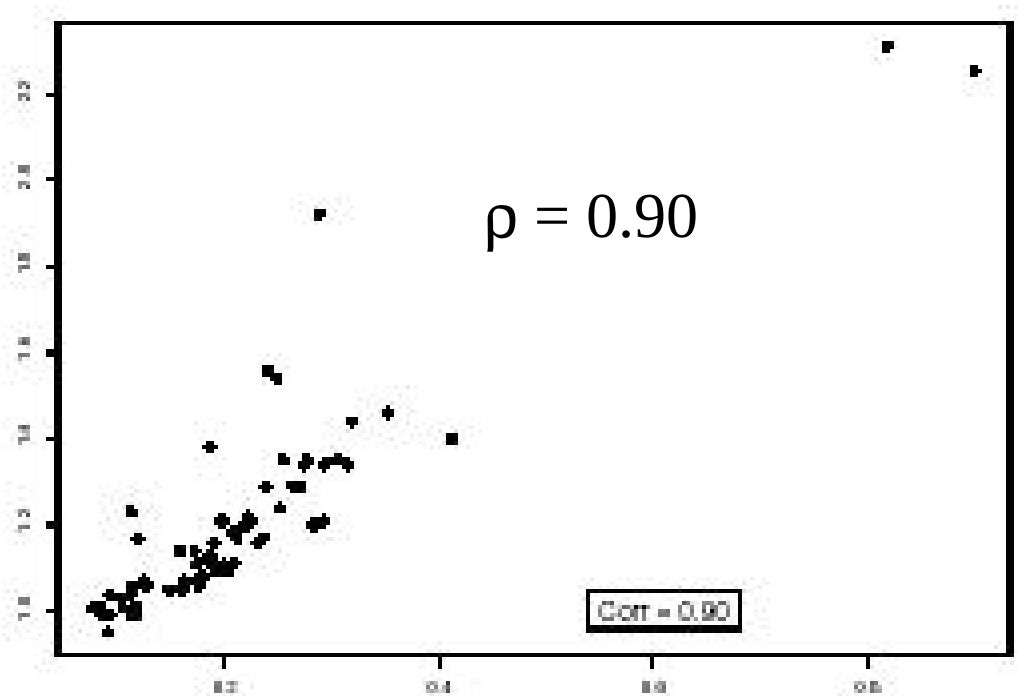
    Others: 98.4%

- Worst performance

  Root: 1.41 sec, gTLD: 0.89sec

- ISPs:
  - Improve performance it provides from gTLD servers for non-cached names

Network Location Relative to Other Servers



x-axis – mean response time of last query/response pair

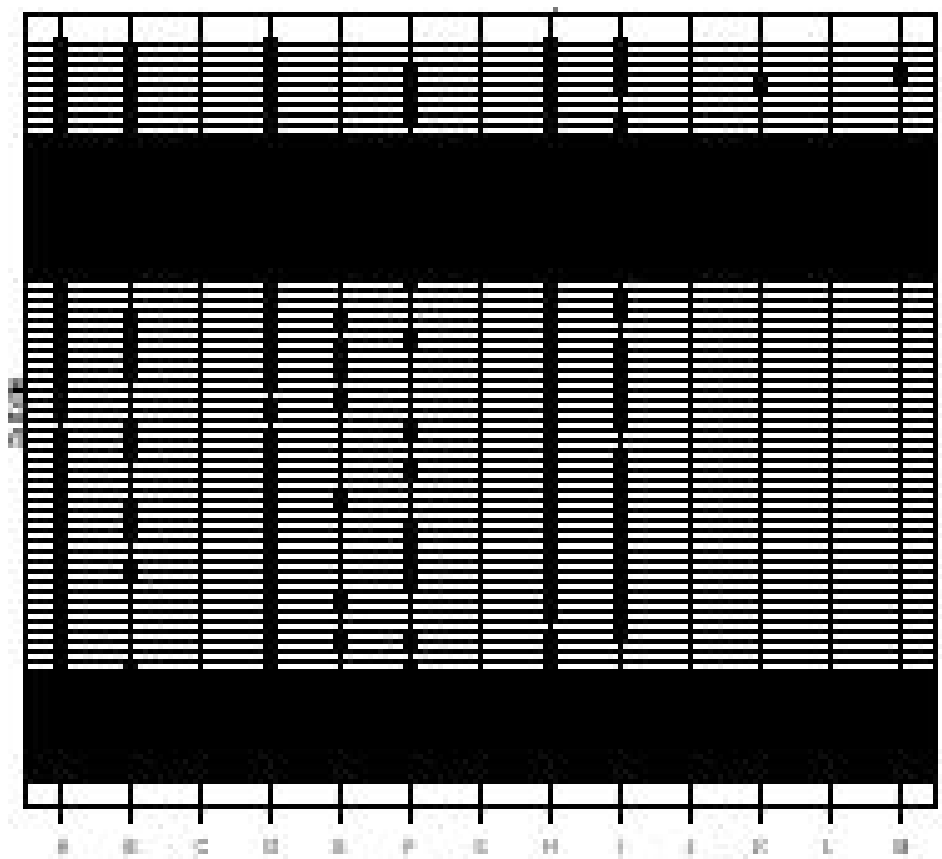
y-axis – mean response time for completed lookups (sec)

- Assumption:
  - Response to the fixed set of servers indicating distance

- Fixed set of servers
  - Last servers queried along the critical path
  - 498 servers w/ same set of IP addresses across all sites

Root Server Interactions

Root servers favored by each site



x-axis – root servers (A – M)

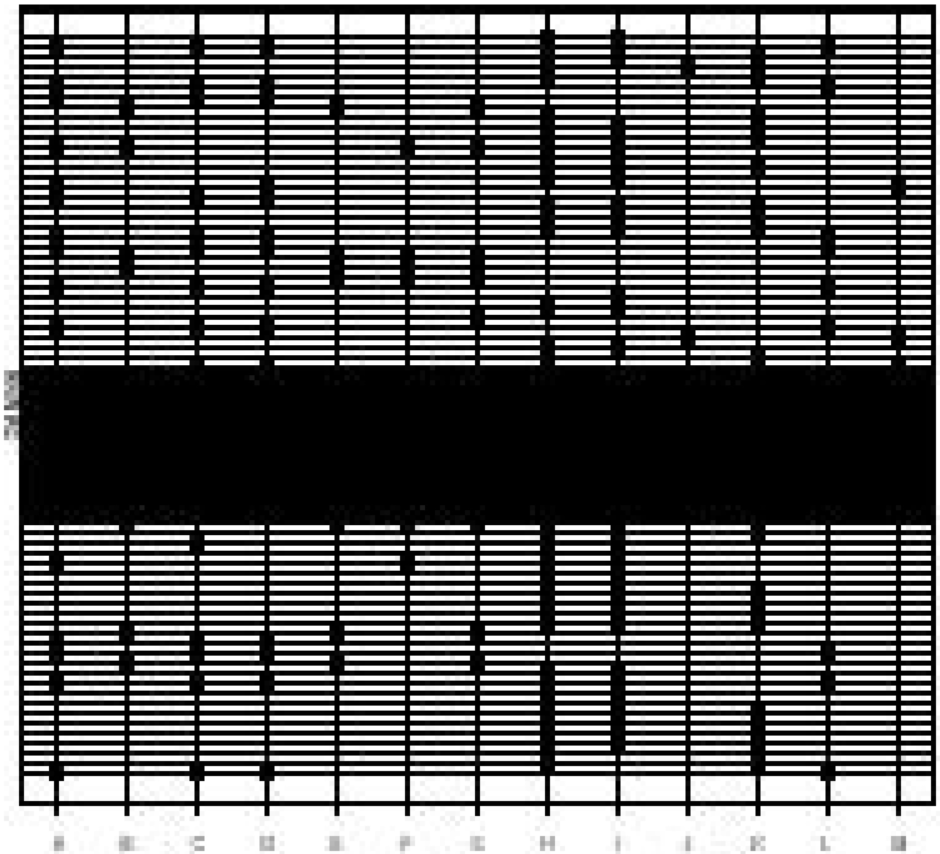
y-axis – sites

A site *favors* a root server if it sends greater than 10% of its root queries to that root server

- Favored by many sites
  - A, D, H, I
- Favored by few or none of the sites
  - C, G, J, K, L, M

gTLD Servers Interaction

gTLD servers favored by each site



x-axis – gTLD servers (A – M)

y-axis – sites

Favored by many sites

– H, I

Favored by few sites

– J, M

- Higher preferences for fewer root servers than gTLD servers
- More variation in favoring gTLD servers from site to site than in root servers

Aliases and CNAMEs

CNAME Redirections

Number of redirections, X	Mean number (percentage) of CNAMEs with X redirections
1	3810 (96.3%)
2	138 (3.5%)
3	8.77 (0.2%)
4	1 (0.03%)

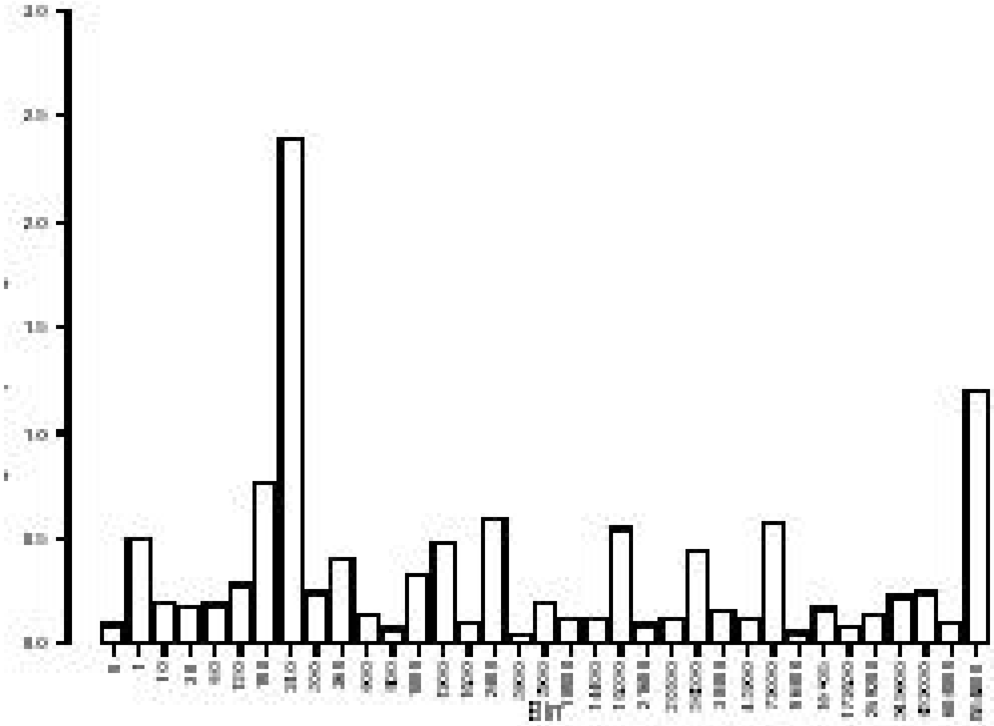
No. of Different CNAMEs per Aliases

Number of different CNAME mappings, X	Number of aliases with X different mappings
1	4230 (93.6%)
2	269 (5.9%)
3	13 (0.2%)
10	1
11	1
15	1
19	1

- About 3960 (26%) of the names in the data set were aliases
- % varied slightly across sites
  - May due to variation of no. of completed lookups
- No. of names that are aliases is not location-sensitive

TTLs of Completed Queries

Ranges of no. of TTLs in each bin across all sites, as a % of the no. of TTLs in the bin



x-axis – bin

y-axis – range as a % of the mean

Bin – chosen based on the modes of the distribution of TTLs for one site

Range – difference b/w the max. no. of TTLs and min. across all sites in each bin

- Extremely small variation in the range of TTLs in each bin
- Distribution of TTLs is invariant across sites

Outline

- "DNS measurements at a root server”

  Nevil Brownlee, Kimberly Claffy, and Evi Nemeth,

  Proceedings of the IEEE GlobeCom, San Antonio, TX, Nov. 2001

- "DNS Performance and the Effectiveness of Caching”

  J. Jung and E. Sit and H. Balakrishnan and R. Morris,

  IEEE/ACM Transactions on Networking, V10, n5, October, 2002

- "Diversity in DNS Performance Measures”

  R. Liston and S. Srinivasan and E. W. Zegura,

  Proceedings Internet Measurement Workshop (IMW), Nov., 2002

- “On the problem of optimization of DNS root servers’ placement”

  Tony Lee, Bradley Huffaker, Marina Fomenkov, kc claffy,

  Passive and Active Measurement Workshop (PAM), La Jolla, CA, 2003

Methodology

Fig. 1. The geographic locations of DNS root servers. Servers marked with '*' currently do not have co-located CAIDA *skitter* monitors. A and J were co-located.

- Data collection – *skitter*
  - Hosts co-located w/ DNS servers
  - Iteratively send 52-byte ICMP echo request packets, incrementally increasing TTL values until a packet reaches the target host
  - Record intermediate router IP addresses and RTT to destination
  - July 14, 2002 – July 20, 2002
  - 3 to 7 RTTs per day for each replying destination

Target List

- Goal
  - Representative
  - One destination in each globally routable prefix from IP addresses sending messages to the DNS root servers
  - 100K-200K addresses
- Tool – *dnsstat*
  - Passively monitor DNS queries at 8 root servers for 24 hours

    A, D, E, F, H, I, K, M
  - 2M client addresses
  - 52K routable prefixes out of 118K prefixes in BGP table from March 18, 2002
- To add destinations uniformly across the IPv4 space
  - Split each /8 prefix into 2 /9 prefix and search for a destination in each half
  - Repeat with next level till /21 level
- Criteria to select among multiple destinations
  - Prefer IP addresses from old DNS Clients list in their previous studies
  - Prefer IP addresses seen by the largest no. of DNS root servers
- 140K destinations

RTT Analysis Assumptions

- Conclusion drawn from the sample of clients are representative of the global DNS system
  - Target list is representative of the overall population of the root servers' clients
- RTT collected by probe ICMP packets are approximately the same as DNS response times actually experienced by root servers' clients
  - Valid only if request processing time < propagation time
- Client selects the best (lowest RTT) available root server
- *Median*(RTT) is a stable and reliable metric of the proximity between two Internet hosts

Significance of Individual Root Servers

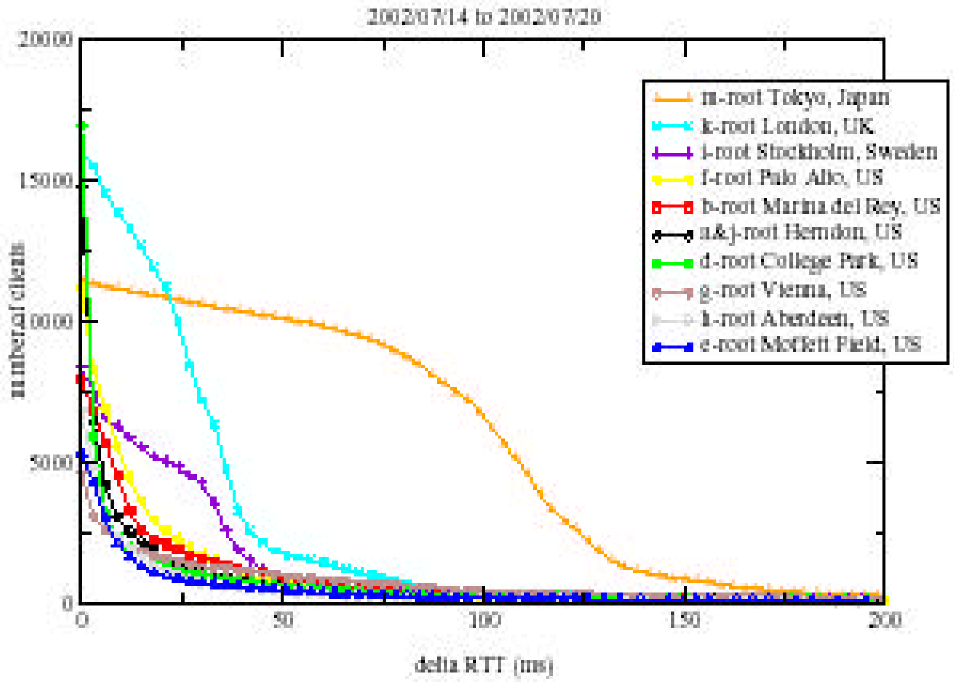


Fig. 2. Increase of latency caused by a root server removal. The curves are CCDFs of the number of clients.

x-axis – closest latency ΔRTT_n

$$\Delta RTT_n = RTT_n^{2nd\_lowest} - RTT_n^{lowest}$$

y-axis – count of clients for which the increase in latency due to removal of their best root server is greater than *x*

- The faster the curve drops along x-axis
  - the fewer clients of this root server are affected
  - the smaller increase in latency clients would experience

- *M* – only root in Asia

  most clients: ≥ 100*ms* increase in latency if removed

- *E* or *H*

  80% clients: < 20*ms* increase in latency if removed

Root Server Clusters

	Group 1 Europe		Group 2 US-East				Group 3 US-West			Group 4 Tokyo, Japan
	k-root	i-root	a&j-root	g-root	h-root	d-root	f-root	e-root	b-root	m-root
k-root	0.0	128.6	191.7	171.0	154.7	151.6	161.1	176.2	194.6	235.7
i-root	128.6	0.0	167.2	174.9	172.4	170.1	181.1	182.7	190.8	232.7
a&j-root	191.7	167.2	0.0	96.5	98.1	97.7	132.6	134.8	141.7	251.9
g-root	171.0	174.9	96.5	0.0	95.5	91.5	128.4	133.5	134.0	231.6
h-root	154.7	172.4	98.1	95.5	0.0	91.5	115.0	120.3	135.9	225.0
d-root	151.6	170.1	97.7	91.5	91.5	0.0	128.3	127.3	138.5	229.3
f-root	161.1	181.1	132.6	128.4	115.0	128.3	0.0	90.2	95.9	196.8
e-root	176.2	182.7	134.8	133.5	120.3	127.3	90.2	0.0	104.2	209.7
b-root	194.6	190.8	141.7	134.0	135.9	138.5	95.9	104.2	0.0	206.1
m-root	235.7	232.7	251.9	231.6	225.0	229.3	196.8	209.7	206.1	0.0

- *distance* between S_1 and S_2

  For each client set $\{client_k\}$, $k=1,\ldots K$

$$D(S_1, S_2) = \frac{1}{K}\sum \left| mRTT_k^{s_1} - mRTT_k^{s_2} \right|$$

- Servers in *Group 1 (Europe)* is less similar to each other than those in *Group 2 and 3 (US)*
  - European servers are geographically more spread out than US servers

Root Server Clusters & Their Clients

Groups	Monitored roots servers	Destinations preferred	All root servers
1. Europe	2 (18.2%)	24,387 (23.7%)	2 (15.4%)
2. US-East	5 (45.5%)	42,978 (41.7%)	6 (46.2%)
3. US-West	3 (27.3%)	24,343 (23.6%)	4 (30.8%)
4. Tokyo, Japan	1 (9.1%)	11,386 (11.0%)	1 (7.7%)
Total	11 (100%)	103,094 (100%)	13 (100%)

Table II. Root families & corresponding subsets of destinations.

Percentages are relative to the total of each column.

- Divide all hosts in the target list into 4 subsets corresponding to 4 groups of root servers
  - Associate a host with a given group if its median RTT is lowest to one of the root servers in this group
- Group 1 (Europe) is most underserved
- US servers are best candidates for relocation to other regions of the world

Impact of a Root Server Relocation

- How would the relocation of existing root servers affect the DNS performance for different groups of clients?

- Use backup server for the *K-root*, *K-peer*, located in Amsterdam
  - Collect 1 week data for all 11 root servers and *K-peer* in July, 2002
  - Suppose *K-peer* replace S_i, service for client n will

    Improve if $mRTT_n^{K\text{-}peer} < mRTT_n^{Si}$

    deteriorate if $mRTT_n^{K\text{-}peer} > mRTT_n^{Si}$

Impact of a Root Server Relocation

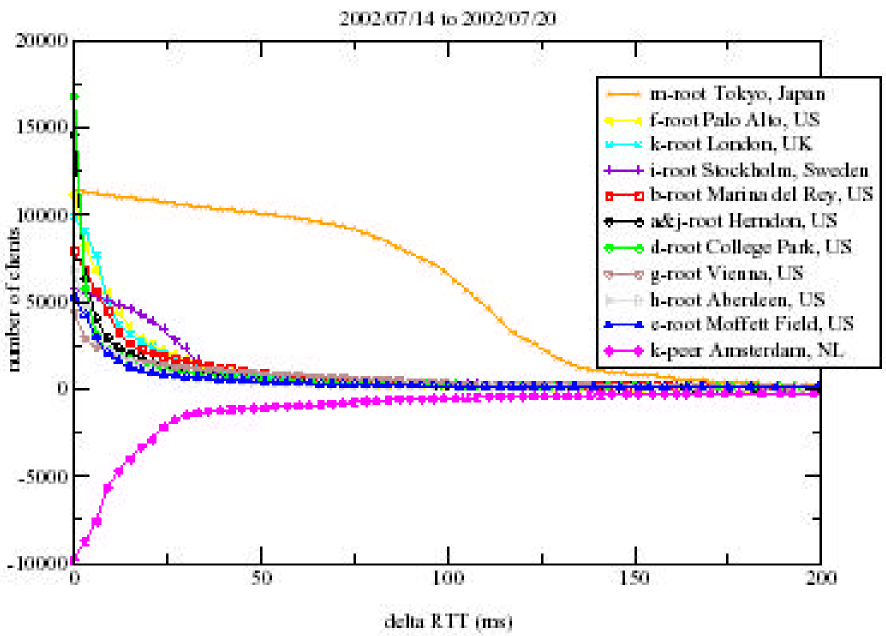


Fig. 4. Latency change caused by a root server relocation.

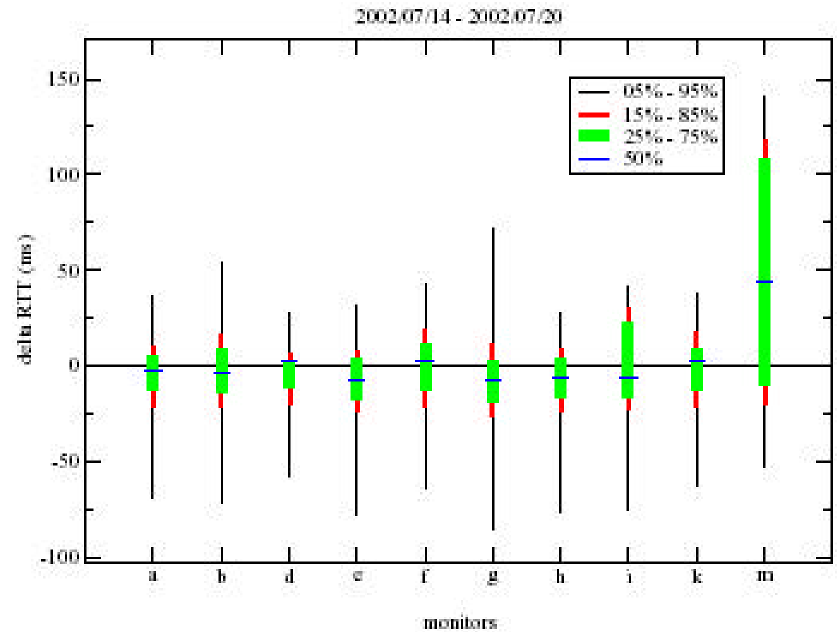


Fig. 5. Combined distributions of latency change due to potential relocation of root name servers.

- Negative curve
  - $mRTT_n^{K\text{-}peer} < mRTT_n^{Si}$ for any $i=1\ldots10$
  - the clients whose connection to the K-peer would have a latency lower than to any root server.

- Root *E*, *G*, and *H* are suitable for relocation
  - the fewest number of clients whose RTT would deteriorate
  - Combined latency distributions are mostly below *x*-axis

?