Interesting DNS Characteristics

Curtis R. Taylor

CS535

December 9, 2013

WPI

Outline

- Motivation and Goal
- Selected Related Work
- Approach
  - Zone files (nameserver data)
    - Offline data linking
  - DNS queries
    - PlanetLab
      - Online and offline collecting and linking
- Results and Discussion
- Lessons learned

Motivation and Goals

- DNS is ubiquitous
  - Performance is important
    - Recall redundancy paper
- Find characteristics about DNS that may lead to interesting research questions or uncover areas for improvement.
  - Track a domain's authoritative nameserver (NS) over a year
  - Perform queries and examine responses

Selected Related Work

- Shaikh *et al.* looked at reduced TTLs values and using DNS to approximate geographical location [1] .
- Cranor *et al.* used DNS traffic from backbone routers to try and identify DNS participants e.g., client, resolver, authoritative servers [2].
- Jung et al. found that reducing TTLs had little adverse effect on cache hit rates [3].
- Shue *et al.* used DNS characteristics to passively link clients to their DNS resolvers [4].

Approach – Tracking (1)

- TLD zone files for over a year (archived each day)
  - (.com, .org, .travel,) .net, .name, .info, etc.
    - Choice based on size
- Information in zone files
  - Domain, authoritative NS domain, IP of NS
    - Not in that order

Approach – Tracking (2)

- Un-archive all zone files
  - How long to track?
    - 1 year (weeks) – Sept. 1, 2012 – Aug. 31, 2013

- Link domain, NS, IP together
  - Reduces number of lookups
  - Cannot link separately (travel needs .com)
  - A script from fellow Grad student will link a given day
  - Time: 4.26 days (12 cores, 64GB RAM)
  - Data: ~44GB/day → 2.4TB total
  - Total domains: ~375 million

Approach – Tracking (3)

- Randomly sort domains and choose 15k from each TLD (45k total) on Sept. 1
- Store domain, NS, and IP
  - Domains may have multiple NS
- For each week, find domain, find NS, check IP
  - Each day's file ~20GB

Approach – Queries (1)

- Choose another random 15k domains from each TLD (45k total)
  - Chosen from Aug. 31, 2013
- For each TLD
  - Ask NS for A record of domain and the following subdomains: www, web, ftp, mail
  - Capture all requests/responses

Approach – Queries (2)

- PlanetLab
  - Obtain slice
  - Add public key
  - Find active nodes
  - Add nodes to slice
  - CoDeploy to distribute software (distributed fashion – ended up dropping)
  - MultiQuery to execute software on machines (ended up dropping)
  - Most nodes at a Univ.

Approach – Queries (3)

- Required 15 nodes
  - 3 TLDs * 5 queries → 4 subdomains + domain
  - 225k queries

- Each node runs a script
  - Install BIND utilities (dig)
  - Download list of domains from remote server
  - Start tcpdump for DNS traffic
  - Issue 15k requests (1 second sleeps)

Approach – Queries (4)

- Each node “phones home” when finished
  - Found many nodes never finished testing scripts from weeks ago. PL is “iffy”
- Copy packet captures locally for analysis
  - Most captures ~3.5MB

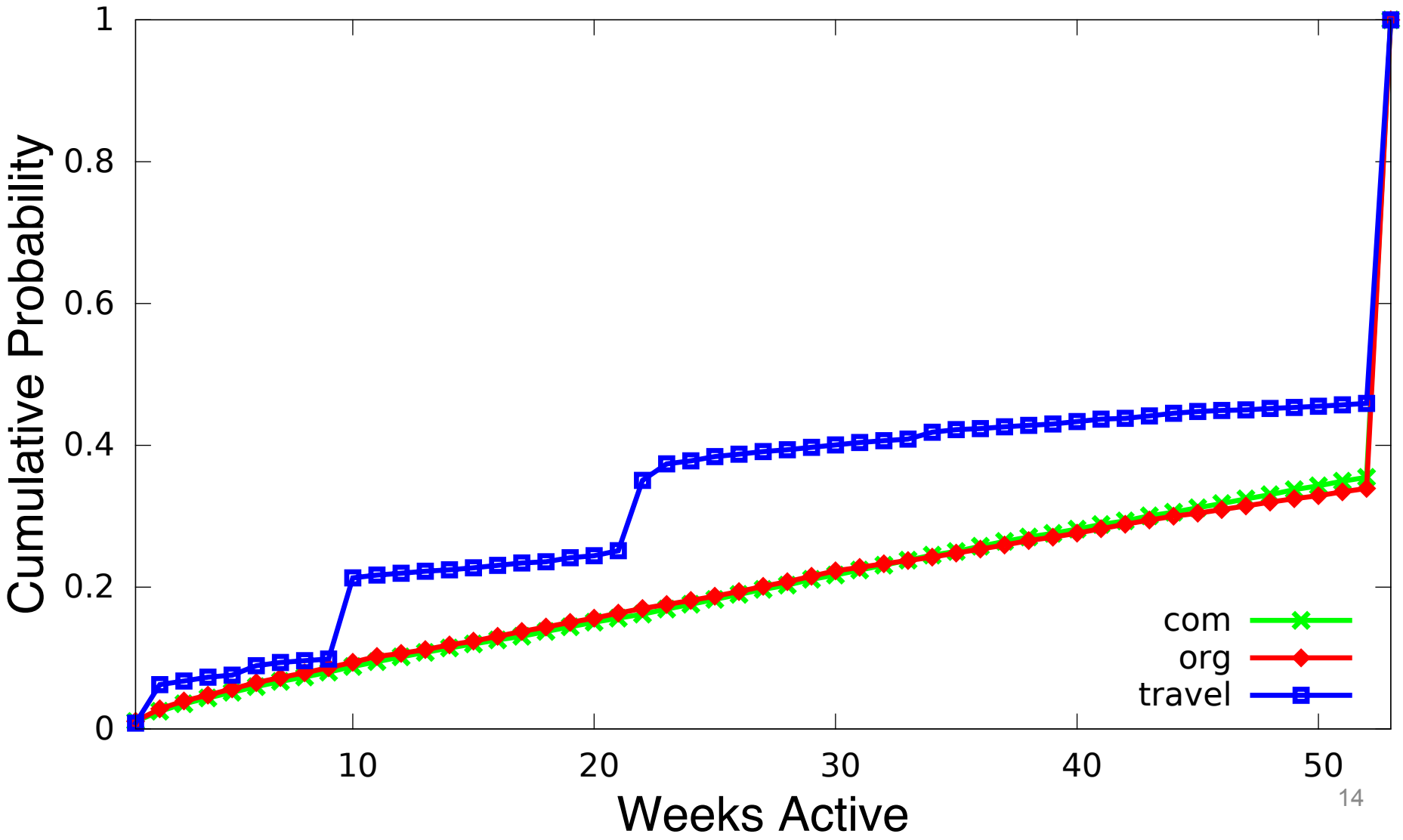
Implementation

- C++ for linking
  - Maps
- Perl scripts for everything else
  - Jim Clausing - SANS Institute
    - Library issues

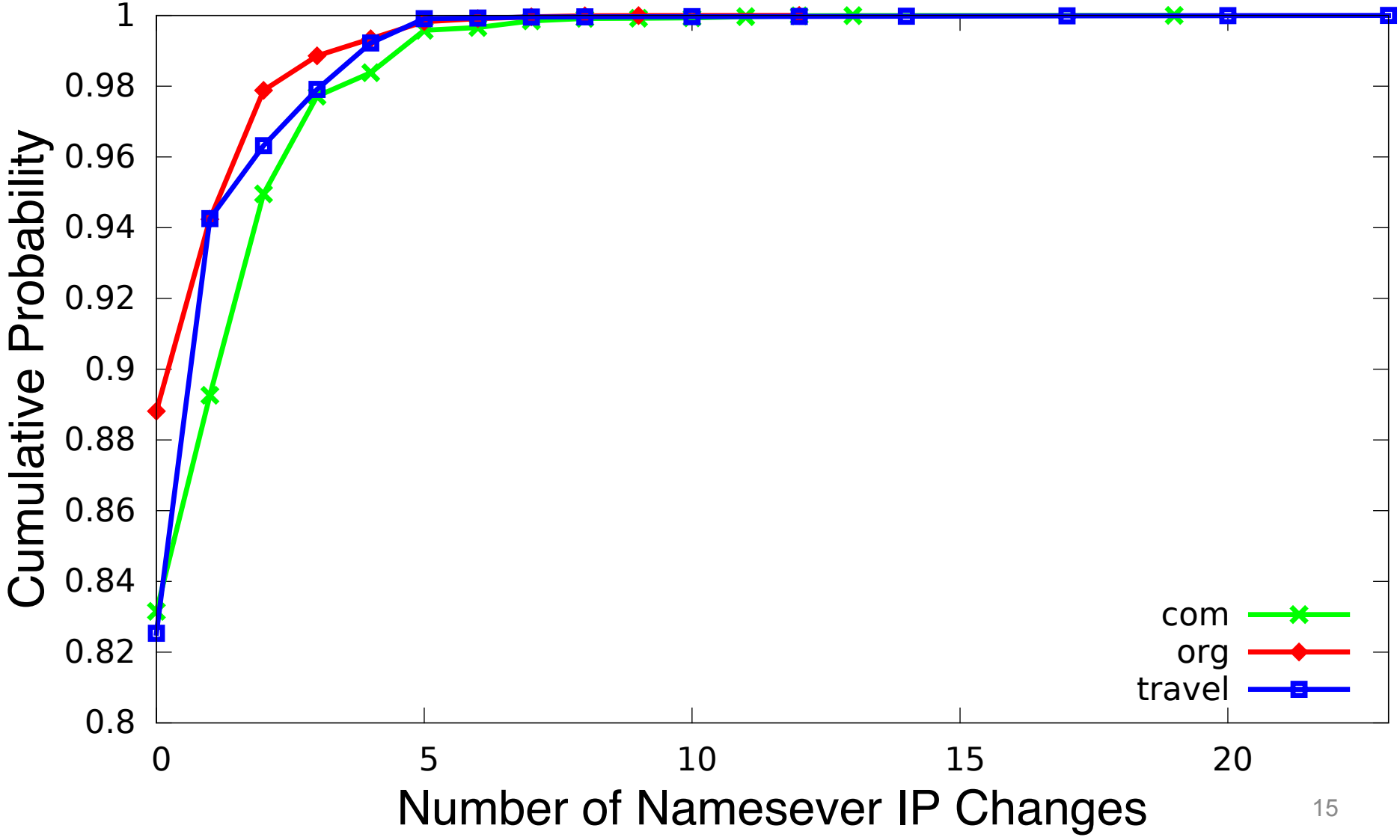
Results – Tracking (1)

- Domains may have many nameservers
  - Made tracking more difficult. Attempted to track which ever NS domain/IP was chosen first. If not found for a given date, considered an NS change

Results – Tracking (2)



Results – Tracking (3)



Results – Queries (1)

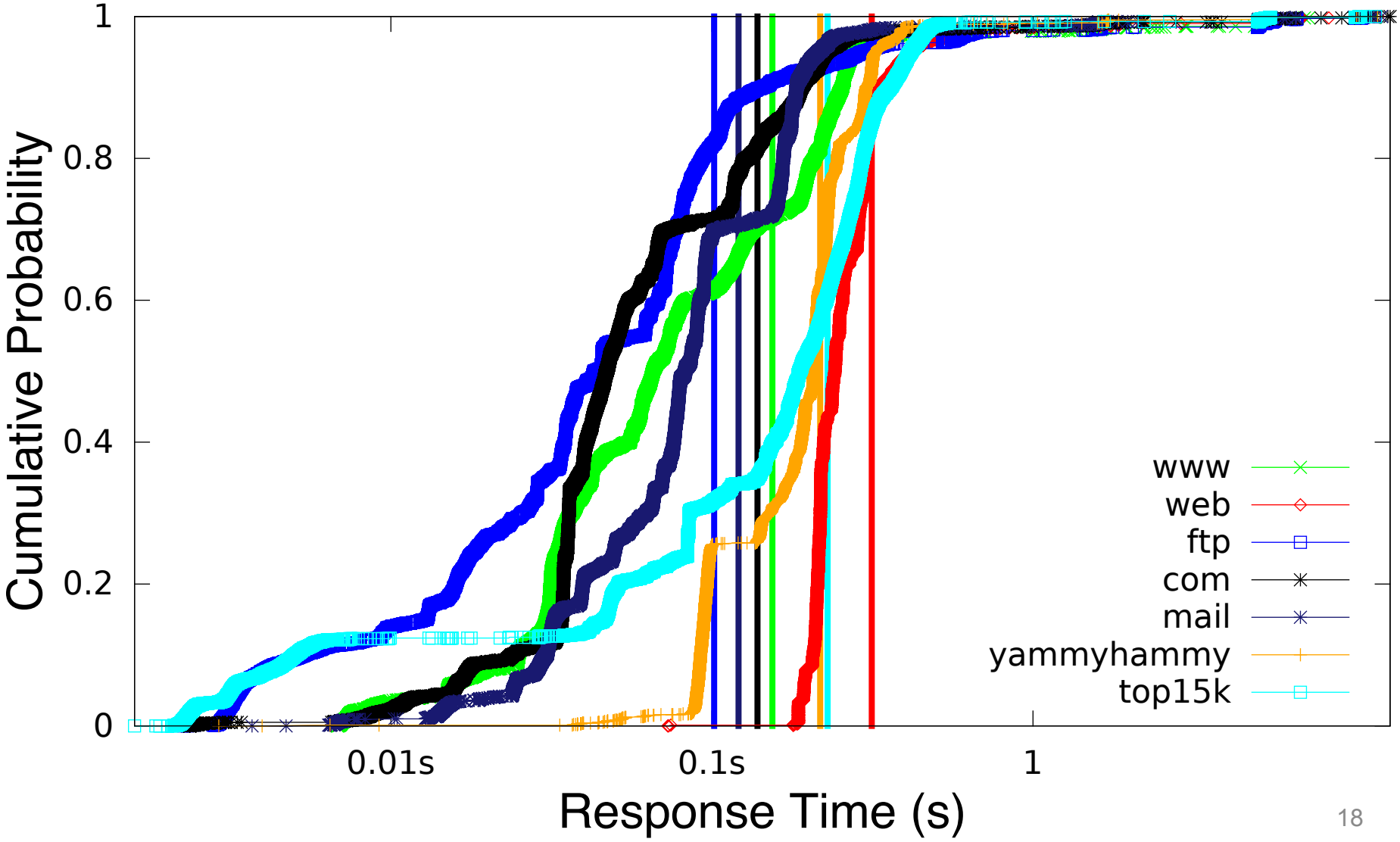
TLD	FQDN	Answers	Avg. A TTL (s)	Avg. CNAME TTL (s)	Avg. Suc. Time (s)
.com	X.com	12282	11111	7438	0.139
.org	X.org	12172	11909	16713	0.186
.travel	X.travel	11110	21175	14416	0.143
.com	www.X.com	11461	11345	10783	0.124
.org	www.X.org	10763	12941	11647	0.154
.travel	www.X.travel	10757	20082	19923	0.106
.com	web.X.com	5159	7737	13066	0.314
.org	web.X.org	4832	7706	14363	0.152
.travel	web.X.travel	2139	18947	27918	0.384
.com	ftp.X.com	8999	10355	10740	0.101
.org	ftp.X.org	9206	10856	11182	0.073
.travel	ftp.X.travel	1391	17015	32888	0.445
.com	mail.X.com	6212	13103	10402	0.121
.org	mail.X.org	5702	15212	10399	0.136
.travel	mail.X.travel	4083	23177	28401	0.179
.com	yammyhammy.X.com	4486	8125	14455	0.217
.com	Top 15K .com	13664	17872	9236	0.229

Results – Queries (2)

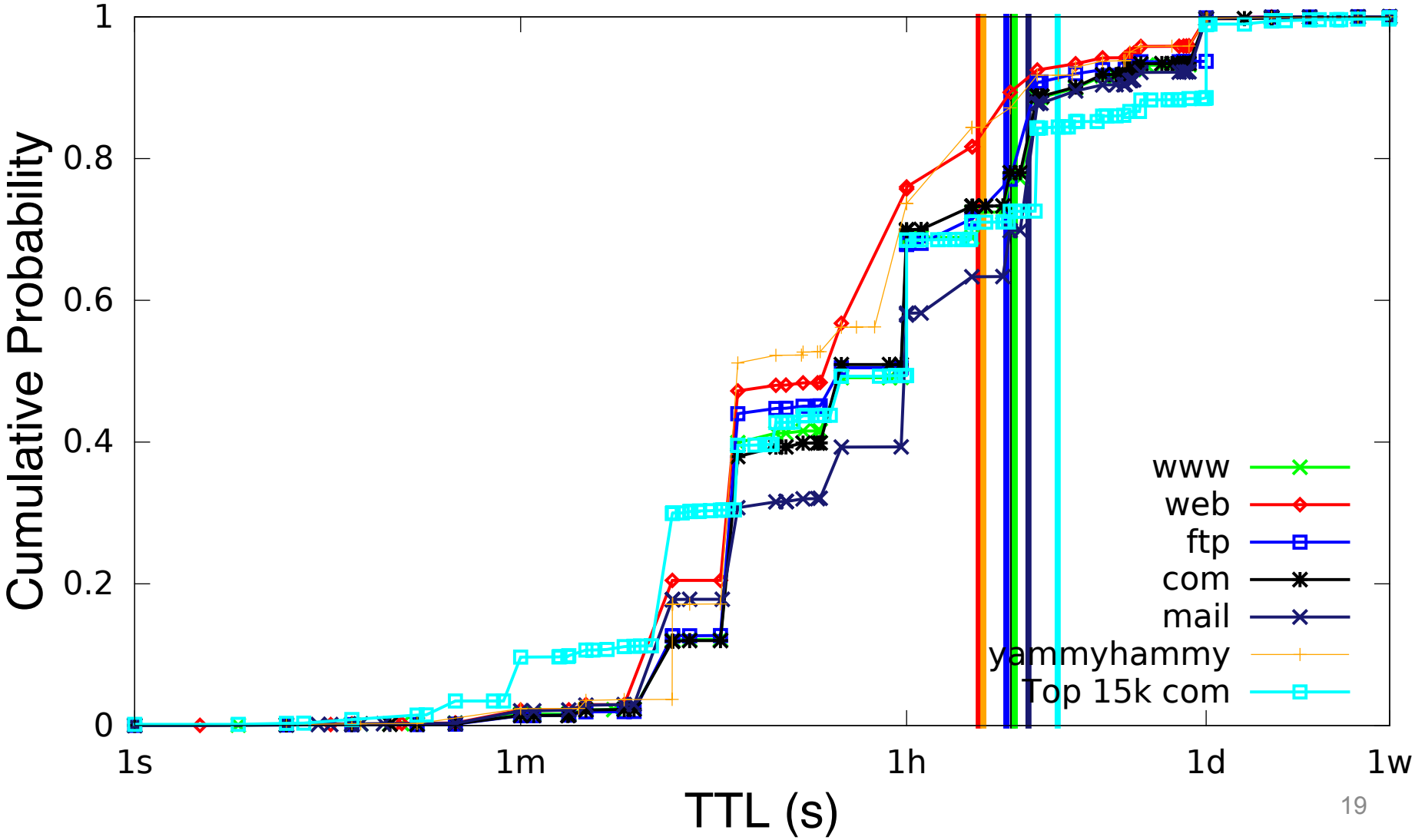
- Based on table wasn't finding distinguishing characteristics

- Two new tests for .com
  - Attempt lookups for random subdomain (yammyhammy.X.com)
  - Top 15k via Alexa (allows top 1m CSV download)
  - Linked top 15k to 20130831 to find NS IP
    - There were top domains not listed in the database

Results – Queries (3)



Results – Queries (4)



Results – Queries (5)

- DNS server are using wildcards
  - web.X.org ~= yammyhammy.X.com
  - web doesn't appear very popular
  - Tested wildcard functionality in BIND9. It might actually have security applications…?
- A few servers required TCP DNS request
  - Usually TCP due to size but were <300 bytes. Far less than UDP max
- Some response times were in the thousands (0.001 place)

Results – Queries (6)

- Found a max TTL in com at 2592000 seconds
  - 2592000s = 30 days
- Found way many more CNAMEs than expected
  - Maybe CNAME was wildcard to main domain
  - CNAME TTL != A record it points to. Problematic?

Lessons Learned (1)

- Dealing with large data sets is different
  - $O(n^2)$ logic in some places – bad
- Had Perl code that would have taken at least 12 hours to run. Implemented with C++ maps and finished in 2 minutes.
- Test thoroughly in small cases.
  - Overnight code crashed due to exception or didn't capture data that I really needed.
  - Over capture and filter afterwards

Lessons Learned (2)

- POP3 or IMAP instead of mail

- Some versions of tcpdump limit the packet capture sizes unless you use “-s” flag.
  - Did captures twice…

Questions?

- [1] A. Shaikh, R. Tewari, and M. Agrawal, “On the effectiveness of DNS-based server selection,” in INFO- COM 2001. Twentieth Annual Joint Conference of the IEEE Computer and Communications Societies. Proceedings. IEEE, vol. 3, 2001, pp. 1801–1810 vol.3.

- [2] C. D. Cranor, E. Gansner, B. Krishnamurthy, and O. Spatscheck, “Characterizing large DNS traces using graphs,” in Proceedings of the 1st ACM SIGCOMM Workshop on Internet Measurement, ser. IMW ’01. New York, NY, USA: ACM, 2001, pp. 55–67. [Online]. Available: http://doi.acm.org/10.1145/505202.505210

- [3] J. Jung, E. Sit, H. Balakrishnan, and R. Morris, “DNS performance and the effectiveness of caching,” IEEE/ACM Trans. Netw., vol. 10, no. 5, pp. 589–603, Oct. 2002. [Online]. Available: http://dx.doi.org/10.1109/TNET.2002.803905

- [4] C. A. Shue and A. J. Kalafut, “Resolvers revealed: Characterizing DNSresolvers and their clients,” ACM Trans. Internet Technol., vol. 12, no. 4, pp. 14:1–14:17, Jul. 2013. [Online]. Available: http://doi.acm.org/10.1145/2499926.2499928