Recommendations for Engineering Authoritative DNS Servers

Giovane Moura[1], Ricardo Schmidt[1,2], Moritz Müller[1,2],
Wouter B. de Vries[2], and John Heidemann[3]

[1]SIDN Labs, [2]University of Twente,
[3]University of Southern California/Information Sciences Institute

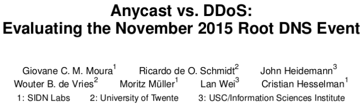
IEPG Meeting @ IETF101
March 18th, 2018
London, UK

Introduction

- *“That’s kind of a vague title”*
- *“What do you mean by “recommendations”?”*

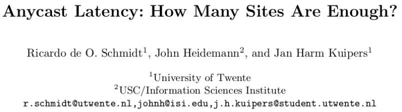
Here we go:

1. Take 4 of our DNS-related papers (3 IMCs, 1 PAM)
2. Summarize their main **take away lessons for operators**



**Anycast vs. DDoS:
Evaluating the November 2015 Root DNS Event**

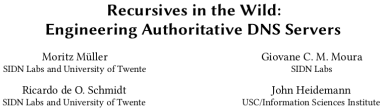
Giovane C. M. Moura[1] Ricardo de O. Schmidt[2] John Heidemann[3]
Wouter B. de Vries[2] Moritz Müller[1] Lan Wei[3] Cristian Hesselman[1]
1: SIDN Labs 2: University of Twente 3: USC/Information Sciences Institute



Anycast Latency: How Many Sites Are Enough?

Ricardo de O. Schmidt[1], John Heidemann[2], and Jan Harm Kuipers[1]

[1]University of Twente
[2]USC/Information Sciences Institute
r.schmidt@utwente.nl,johnh@isi.edu,j.h.kuipers@student.utwente.nl



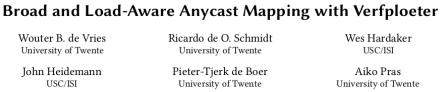
**Recursives in the Wild:
Engineering Authoritative DNS Servers**

Moritz Müller
SIDN Labs and University of Twente

Giovane C. M. Moura
SIDN Labs

Ricardo de O. Schmidt
SIDN Labs and University of Twente

John Heidemann
USC/Information Sciences Institute



Broad and Load-Aware Anycast Mapping with Verfploeter

Wouter B. de Vries
University of Twente

Ricardo de O. Schmidt
University of Twente

Wes Hardaker
USC/ISI

John Heidemann
USC/ISI

Pieter-Tjerk de Boer
University of Twente

Aiko Pras
University of Twente

Recommendations

- R1: all authoritatives should have similar latency [1]
- R2: Routing Can Matter More Than Locations [2]
- R3: Detailed Anycast Maps of Catchments Requires Active Measurements [3]
- R4: When under stress, two strategies[4]
- R5: Shared Infrastructure Risks Collateral Damage During Attacks [4]

R1: all authoritatives should have similar latency

- DNS operators run their zones on multiple authoritative servers
  - NS records
- Each of them may use anycast
  - 13 NSes for the roots, 1000s of servers
- Operators strive to reduce latency for users
- But they only control part of the infrastructure
- And not how the recursives (user side) will choose authoritatives

R1: all authoritatives should have similar latency

- We set to answer how **recursives choose authoritatives in the wild**
- We set up 7 NSes (1 per EC2 area)
- Then, we ran the same DNS zone with various NS setups:
  - Varying **number of NSes**: 2, 3 and 4
  - Varying **locations**: FRA, DUB, IAD, SFO, GRU, NRT , SYD
- Used 10,000 Ripe Atlas probes as vantage points (VPs)
- Analyze how VPs' recursives choose from available NSes

R1: all authoritatives should have similar latency

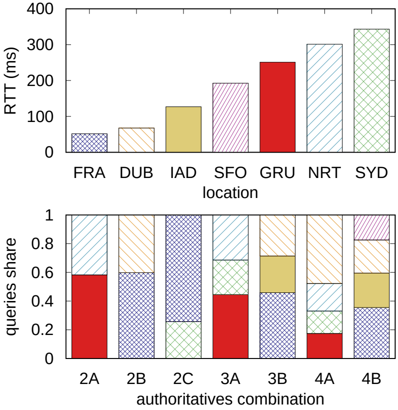


Figure: Query distribution (top) and median RTT (bottom) for combinations of authoritatives.

R1: all authoritatives should have similar latency

- Our hypothesis: recursives use performance (lower latency) and diversity of NSes when choosing
- For a DNS operator, this policy means that *latency of all authoritatives matter, so all must be similarily capable*, since all available authoritatives will be queried by most recursives.
- Since IP unicast cannot deliver good latency worldwide, we recommend operators to deploy equally strong IP anycast in every NS.
    - That's what are doing at `.nl`

R2: Routing can matter more than locations

- Say you want to hire a DNS provider
- Which criteria would you employ, besides pricing?
- Number of anycast sites is often a chosen metric
  - The more the merrier?
  - Meaning you have more servers distributed across the globe, therefore serving better your users
- We found that *this is not necessarily true*
- Actually, routing can matter more than number of sites/locations

R2: Routing can matter more than locations

- We analyzed the relationship between number of anycast sites and RTT for:
  - C,F,K and L root
  - Using 7.9K Ripe Atlas probes (VPs)

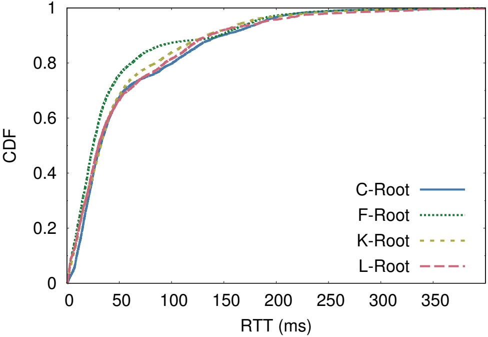
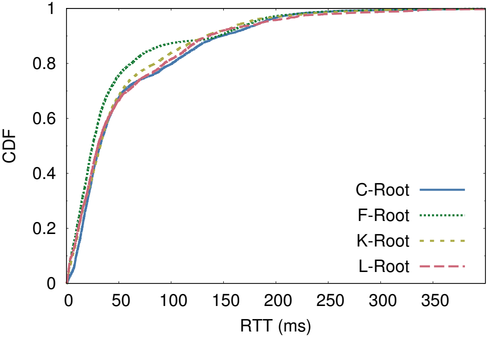


Figure: CDF of observed latency for C, F, K and L-Root servers.

R2: Routing can matter more than locations



- C-Root (8 sites at the time) had similar performance (RTT) to larger services:
  - K (33 sites), L(144 Sites)
  - C, K, and L: RTT between 30 and 30ms
  - F Rooot: 25ms

R2: Routing can matter more than locations

- Not in the study: one DNS provider with 80+ sites (including SFO) answers its DNS queries from Amazon EC2 Northern California from Tokyo instead!
- Peering between both is the issue
- So our recommendation: consider also the location of the sites when choosing a DNS provider
- Closest to your users (in BGP terms, not only geo)
- More sites, however, can provide extra resilience under a DDoS attack

R3: Detailed Anycast Maps of Catchments Requires Active Measurements

- Say you run a 20 site anycast DNS service
- BGP will match your users to their “nearby” site:
  - Nearby in terms of BGP routing
- Adding an extra site may change entirely the load distribution across your sites:
  - And suddenly your have underused and overload sites
- So it’s very trick to predict how the traffic will shift after adding sites

R3: Detailed Anycast Maps of Catchments Requires Active Measurements

- To handle that, we developed Verfploeter:
  - An open source tool/technique that can be used by operators to predict catchment (where BGP will send users) and query load
- We used to predict catchment shifts on B-root (2 sites) :
  - We estimated 81.6% of the traffic would go to LAX
  - And in practice, 81.4% did go
- How it works?
  1. Create catchment maps: send ICMP packets to every /24 on anycast address, than see in which site the echo replies end
  2. Use this map to estimate your traffic load by:
     - Looking at your current traffic distribution
     - Matching it with the mappings

R3: Detailed Anycast Maps of Catchments Requires Active Measurements

- It can also be used to estimate traffic shift during a DDoS
- Like, if you prepend routes, what happens with the traffic?

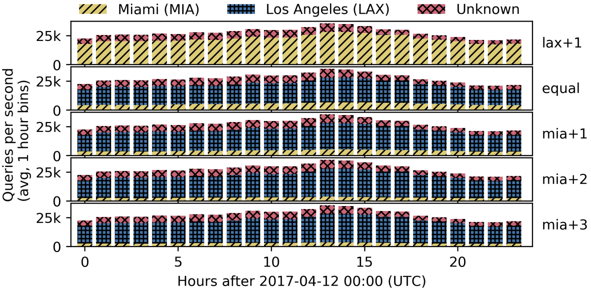


Figure: Load on new B-root deployment during a day, using production logs from the previous unicast setup. +n indicates AS Path prepending.

R3: Detailed Anycast Maps of Catchments Requires Active Measurements

- Our recommendation for DNS operators is:
  - If you expand or engineer a new service, use Verflploeter to make informed choices on how engineer your service
  - Open-source tool

R4: When under stress, two strategies

- DDoS are becoming bigger and cheaper
- 1.2Tb/s is the current record; not sign of going away soon
- So what do do under stress for your Anycast NS?
- We investigated this question using empirical observations from the Root DNS events of Nov 30th, 2015
    - 35 Gb/s direct attack of legitimate DNS queries

R4: When under stress, two strategies

So what are the strategies?

1. Try to **redirect traffic** with withdraw/prepending routes
   - That will cause the catchment to shrink and shift traffic to bigger sites (Verflploeter can estimate where exacly)
2. Or you can "**sacrifice**" one or few sites
   - You man want to leave one site to absorb most of the attack
   - So users elsewhere can have normal services

- We saw both during the DDoS against the roots
- And we need to investigate more careful and informed choices
  - We have a new project coming up for that

R5: Shared Infrastructure Risks Collateral Damage During Attacks

- So when you hire a DNS provider, you'll share some infrastructure
- There are pros and cons of that:
  - May be cheaper
  - Bigger infrastructure than you'd have
  - Diversity
- However, things may get ugly during a DDoS
  - If one zone is target, all the others they share may have trouble
- We have seen it with the 1.2Tb/s Mirai attack: many clients of the DNS provider suffered

R5: Shared Infrastructure Risks Collateral Damage During Attacks

- Collateral damage during the Root DNS event
- D-ROOT **was not** attacked!

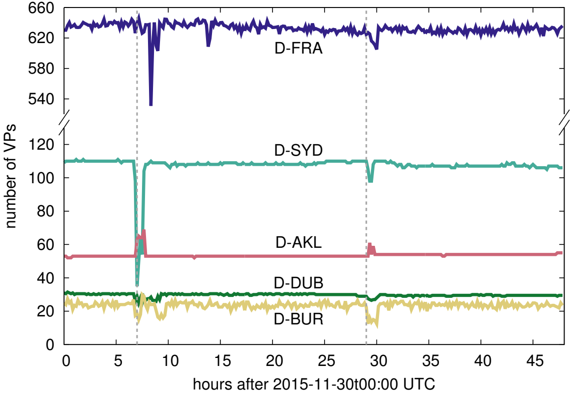


Figure: Reachability of those D-Root sites that were affected by the DDoS.

R5: Shared Infrastructure Risks Collateral Damage During Attacks

- Collateral damage during the Root DNS event
- Neither .nl was attacked

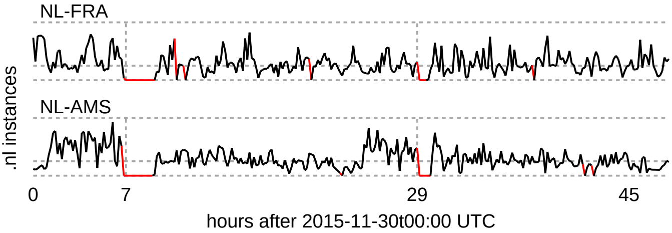


Figure: Normalized number of queries for .nl , measured at the servers in 10 min bins.

R5: Shared Infrastructure Risks Collateral Damage During Attacks

- Our recommendation for operators is: be aware of shared infrastructure
- It may increase the attack surface during a DDoS

Summary

Recommendations for operators from 4 of our papers:

- R1: all authoritatives should have similar latency [1]
- R2: Routing Can Matter More Than Locations [2]
- R3: Detailed Anycast Maps of Catchments Requires Active Measurements [3]
- R4: When under stress, two strategies[4]
- R5: Shared Infrastructure Risks Collateral Damage During Attacks [4]

Bibliography I

[1] M. Müller, G. C. M. Moura, R. de O. Schmidt, and J. Heidemann, "Recursives in the wild: Engineering authoritative DNS servers," in *Proceedings of the ACM Internet Measurement Conference*, London, UK, 2017. [Online]. Available: http://www.isi.edu/%7ejohnh/PAPERS/Mueller17b.html

[2] R. d. O. Schmidt, J. Heidemann, and J. H. Kuipers, "Anycast latency: How many sites are enough?" in *Proceedings of the Passive and Active Measurement Workshop*. Sydney, Australia: Springer, Mar. 2017, pp. 188–200. [Online]. Available: http://www.isi.edu/%7ejohnh/PAPERS/Schmidt17a.html

Bibliography II

[3] W. B. de Vries, R. de O. Schmidt, W. Haraker, J. Heidemann, P.-T. de Boer, and A. Pras, “Verfploeter: Broad and load-aware anycast mapping,” in *Proceedings of the ACM Internet Measurement Conference*, London, UK, 2017. [Online]. Available: http://www.isi.edu/%7ejohnh/PAPERS/Vries17b.html

[4] G. C. M. Moura, R. de O. Schmidt, J. Heidemann, W. B. de Vries, M. Müller, L. Wei, and C. Hesselman, “Anycast vs. DDoS: Evaluating the November 2015 Root DNS Event,” in *Proceedings of the ACM Internet Measurement Conference*, Nov. 2016. [Online]. Available: http://www.isi.edu/%7ejohnh/PAPERS/Moura16b.html

Questions?

- giovane.moura@sidn.nl