The Internet of Names: a DNS Big Dataset

Actively Measuring 50% of the Entire DNS Name Space, Every Day

Roland van Rijswijk-Deij*† r.m.vanrijswijk@utwente.nl

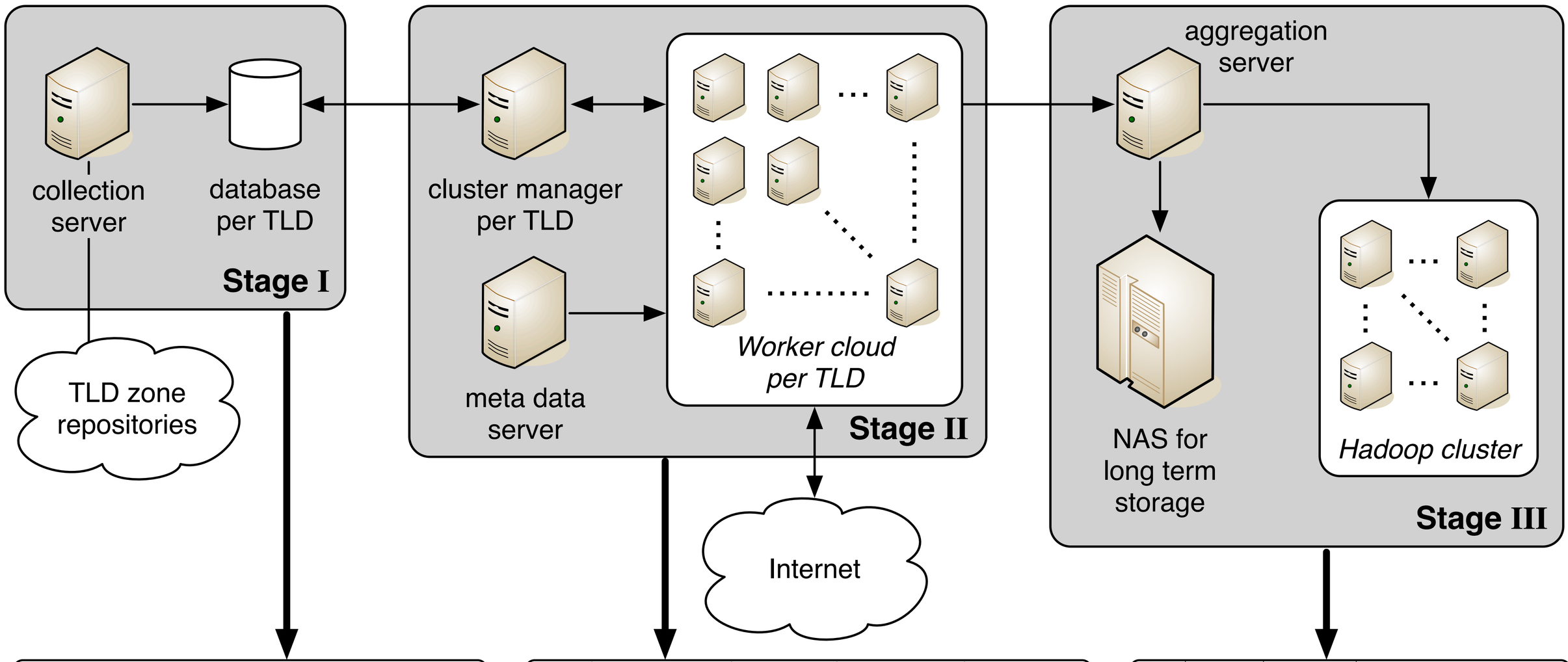
Mattijs Jonker* m.jonker@utwente.nl

Anna Sperotto* a.sperotto@utwente.nl

Aiko Pras* a.pras@utwente.nl

* Design and Analysis of Communication Systems (DACS), Faculty of Electrical Engineering, Mathematics and Computer Science, University of Twente, Enschede, The Netherlands

† SURFnet bv, Utrecht, The Netherlands



TLD	#domains	(% of DNS)	Size	Stage I run-time
.com	117.8M	(40.1%)	2345MB	4h 32 min.
.net	14.9M	(5.1%)	306MB	40 min.
.org	10.5M	(3.6%)	183MB	17 min.
total	143.2M	(48.7%)	2834MB	5h 29 min.

Table 1: Input zones

TLD	#worker VMs	#batches	time (batch)	time (total)
.com	80	1180	59 min.	17h 30 min.
.net	10	149	55 min.	14h 30 min.
.org	10	106	41 min.	7h 45 min.

Table 2: Measurement infrastructure

TLD	#results	#domains	size	(uncompressed)
.com	1374M	116.9M	27.9GB	(204.7GB)
.net	163M	14.7M	3.3GB	(23.9GB)
.org	122M	10.4M	2.5GB	(18.0GB)
total	1658M	141.9M	33.6GB	(246.7GB)

Table 3: Sample results for one day

Why Measure DNS?

The Domain Name System is used by almost every networked service. It maps human readable names to IP addresses, but also, for instance, which hosts handle e-mail for a domain or information about PKI certificates. Thus, measuring what is in the DNS can tell us a lot about the state of the Internet.

Our Goals:

We want to measure the DNS to track the evolution of the Internet over time. We therefore want to:

- Measure every domain in the main top-level domains
- Collect data for each domain at least once every 24 hours
- Store at least 1 year of data

Query Types

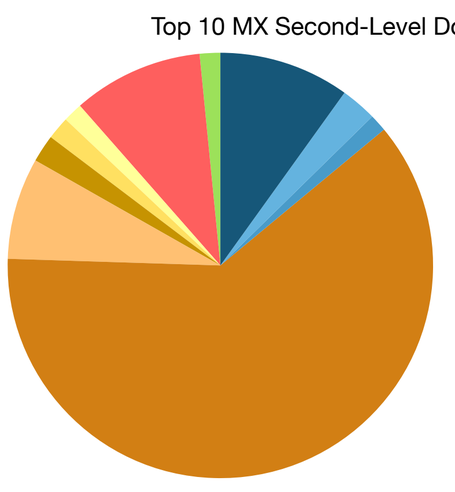
- **SOA** DNS zone configuration
- **A** IPv4 address (for apex, 'www' and 'mail')
- **AAAA** IPv6 address (for apex, 'www' and 'mail')
- **NS** Authoritative name servers
- **MX** Mail eXchangers
- **TXT** Arbitrary text, used for SPF, DKIM, DMARC, proof of domain ownership, …
- **DS** Delegation Signer (secure DNSSEC delegation)
- **DNSKEY** Public keys used in DNSSEC signing
- **NSEC(3)** Authenticated Denial-of-Existence (used by DNSSEC)

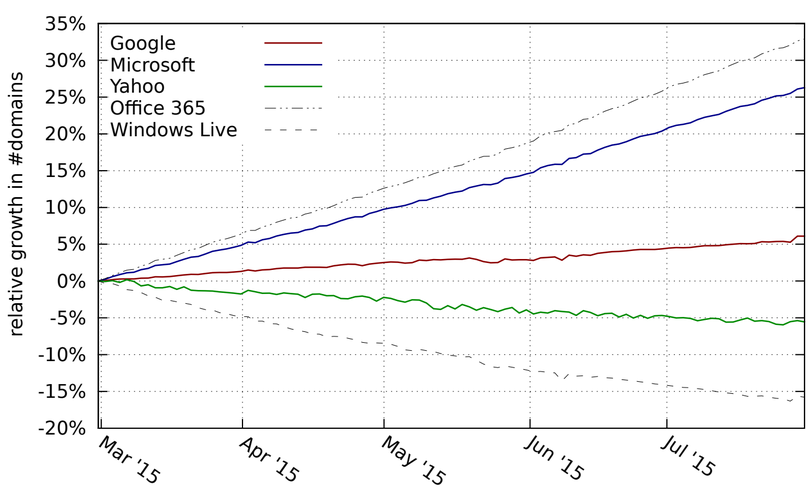
Data Sharing

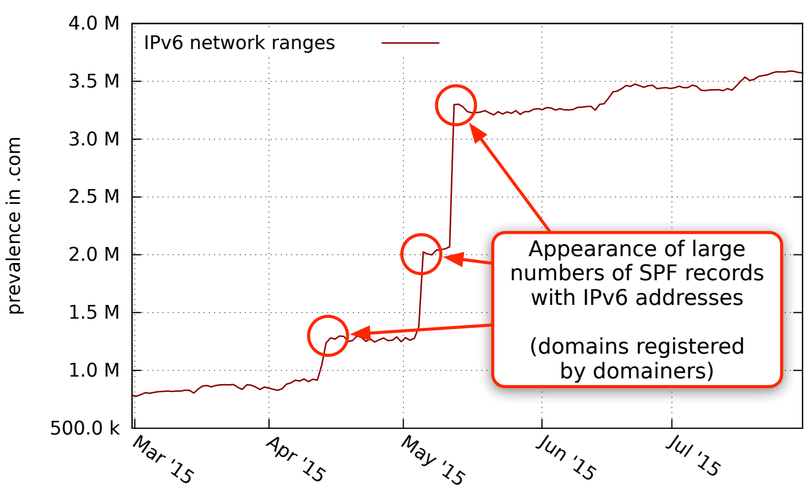
We are working on a public web portal to share aggregate datasets (e.g. number of domains with at least one *AAAA* record per country, …). Next to that, we are setting up a programme for researchers to visit our group to use the data we collect (which cannot be made public due to contractual constraints). **If you are interested in visiting us, please contact us via e-mail!**

Future Work

- Extend the measurement with additional (cc)TLDs
- Collaborate with CSIRT teams to explore security applications of the data







Example 1: Top Mail Handlers in .com

Takeaways:
- Hosters/registrars dominant
- Cloud e-mail services clearly visible
- "Domainers" significant presence

Example 2: Growth of Cloud E-mail

Takeaways:
- Use of cloud e-mail is growing
- Microsoft Office 365 fastest grower
- Windows Live (Hotmail) dying off

Example 3: Anomalies for IPv6 in SPF

Takeaways:
- Dataset well-suited for time series
- Anomalies lead to discoveries
- Sparks new research avenues

DACS Design and Analysis of Communication Systems | UNIVERSITY OF TWENTE. | SURFNET