Academic collaboration with industry: solving real-world networking problems

Giovane C. M.Moura

SIDN Labs and TU Delft

INSY-EWI Faculty Lunch

Delft, The Netherlands

2023-12-07

Stereotypes

Academics seen by industry

Industry seen by academics

Stereotypes

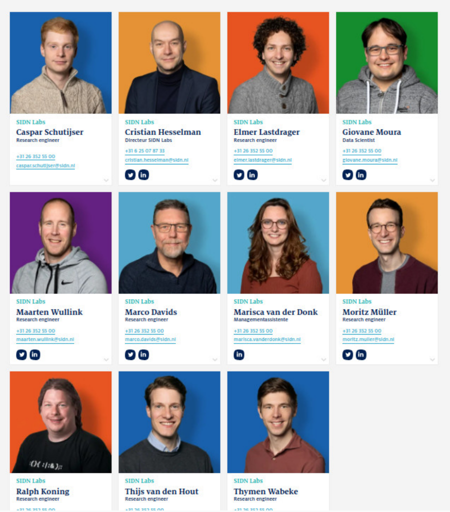
Academics seen by industry

Industry seen by academics

What if we could have a win-win situation?

SIDN Labs

- Research arm of SIDN
  - Bridge between industry and academia
  - 11 ppl; 5 with PhDs, 1 Prof., 1 Assistant Prof., Engineers, and Ops
- (We don't sell anything)
- Three main areas:
  - DNS security
  - Infrastructure security
  - Future Internet
- Deliverables: academic papers, systems, software, standardization



SIDN Labs
Caspar Schutijser
Research engineer
+31 26 352 55 00
caspar.schutijser@sidn.nl

SIDN Labs
Cristian Hesselman
Directeur SIDN Labs
+31 6 25 07 87 33
cristian.hesselman@sidn.nl

SIDN Labs
Elmer Lastdrager
Research engineer
+31 26 352 55 00
elmer.lastdrager@sidn.nl

SIDN Labs
Giovane Moura
Data Scientist
+31 26 352 55 00
giovane.moura@sidn.nl

SIDN Labs
Maarten Wullink
Research engineer
+31 26 352 55 00
maarten.wullink@sidn.nl

SIDN Labs
Marco Davids
Research engineer
+31 26 352 55 00
marco.davids@sidn.nl

SIDN Labs
Marisca van der Donk
Managementassistente
+31 26 352 55 00
marisca.vanderdonk@sidn.nl

SIDN Labs
Moritz Müller
Research engineer
+31 26 352 55 00
moritz.muller@sidn.nl

SIDN Labs
Ralph Koning
Research engineer
+31 26 352 55 00

SIDN Labs
Thijs van den Hout
Research engineer
+31 26 352 55 00

SIDN Labs
Thymen Wabeke
Research engineer
+31 26 352 55 00

What does SIDN do

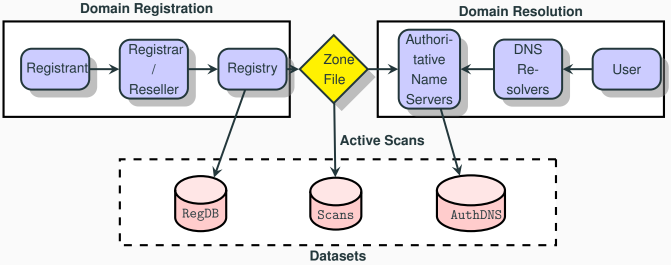
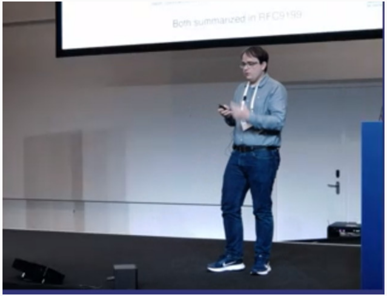


Figure 1: TLD operations: registration (left), domain resolution (right), and datasets.

$whoami

- Data Scientist at SIDN Labs
- Assistant Professor at Cybersecurity group
    - Secondment 1 day a week (*detachering*)
- PhD (2013, UTwente, NL)
- MSc (2008, UFRGS, BR)



Presentation @ RIPE86, Rotterdam, May 2023

Today's presentation

	Academia	**Industry**
Data	some	tons
Money	some	depends
Research skills	tons	little
Time for research	some	barely none

- Three cases of successful industry-academic collaboration

Outline

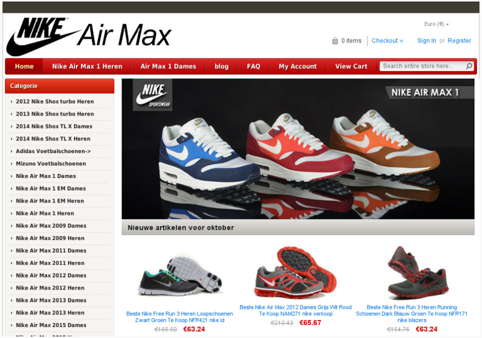
Case 1: Web Security

Case 2: Large Authoritative DNS servers Ops

Case 3: Time services on the Internet

Case #1: Applying academic skills to industry

- We stumbled on these websites while looking for phishing
- They were rather *odd*
- We had many questions:
  1. does anyone even *buy* from them?
  2. what is their *business model*?
  3. how many they were (on .nl)?
  4. what can we do about it?



Categorie

- 2012 Nike Shox turbo Heren
- 2013 Nike Shox turbo Heren
- 2014 Nike Shox TL X Dames
- 2014 Nike Shox TL X Heren
- Adidas Voetbalschoenen->
- Mizuno Voetbalschoenen
- Nike Air Max 1 Dames
- Nike Air Max 1 EM Dames
- Nike Air Max 1 EM Heren
- Nike Air Max 1 Heren
- Nike Air Max 2009 Dames
- Nike Air Max 2009 Heren
- Nike Air Max 2011 Dames
- Nike Air Max 2011 Heren
- Nike Air Max 2012 Dames
- Nike Air Max 2012 Heren
- Nike Air Max 2013 Dames
- Nike Air Max 2013 Heren
- Nike Air Max 2015 Dames

Nieuwe artikelen voor oktober

Beste Nike Free Run 3 Heren Loopschoenen Zwart Groen Te Koop NFR421 nike id
~~€165.50~~ €63.24

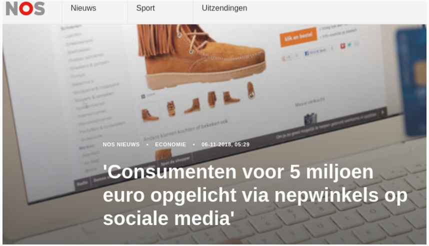
Beste Nike Air Max 2012 Dames Grijs Wit Rood Te Koop NAM271 nike verkoop
~~€210.43~~ €65.67

Beste Nike Free Run 3 Heren Running Schoenen Dark Blauw Groen Te Koop NFR171 nike blazers
~~€154.76~~ €63.24

Figure 2: Screenshot of 2016 `.nl` website

Does anyone even buy from them?

- Yes, they were
- Scam: getting fake or no product
- Dealing with financial losses



NOS NIEUWS • ECONOMIE • 06-11-2018, 05:29

'Consumenten voor 5 miljoen euro opgelicht via nepwinkels op sociale media'

Figure 3: NOS news (2018)

OK, so what to do about it

- SIDN is a Internet registry, not police
- But we have a mission to make the .nl zone safer for users
- And we were sitting on the data

- Ethical dilemma:
  - Turn the blind eye OR
  - Do something about it
- We talked to our lawyers
- We need to conform to our mandate and EU and NL laws

We decided to go ahead and measure it

OK, so what to do about it

- SIDN is a Internet registry, not police
- But we have a mission to make the .nl zone safer for users
- And we were sitting on the data

- Ethical dilemma:
  - Turn the blind eye OR
  - Do something about it
- We talked to our lawyers
- We need to conform to our mandate and EU and NL laws

We decided to go ahead and measure it

OK, so what to do about it

- SIDN is a Internet registry, not police
- But we have a mission to make the .nl zone safer for users
- And we were sitting on the data

- Ethical dilemma:
  - Turn the blind eye OR
  - Do something about it
- We talked to our lawyers
- We need to conform to our mandate and EU and NL laws

We decided to go ahead and measure it

What is their *business model*?

- The business model goes like this:
  1. Consumer demand [3]
  2. Manufacturing in China [1]
  3. These webshops connect both of them
- It's not only a .nl problem:
  - .de, .be, .com, and many others have the same issue
- We are dealing with *pros* here

How many were on the .nl zone?

- We realized they all share a similar pattern:
  1. long `html <title>` tags

```
<title>Vans Schoenen On Sale 70% OFF |Geen
    verzendkosten</title>
```

  2. tags listing many brands (Nike, Reebok, Gucci, you name it..)
- **Question: Why this tactic?**
  - Search Engine optimization → more clicks, more money [4]

How many were on the .nl zone?

- We realized they all share a similar pattern:
  1. long `html <title>` tags

     ```
     <title>Vans Schoenen On Sale 70% OFF |Geen verzendkosten</title>
     ```

  2. tags listing many brands (Nike, Reebok, Gucci, you name it..)
- **Question: Why this tactic?**
  - Search Engine optimization → more clicks, more money [4]

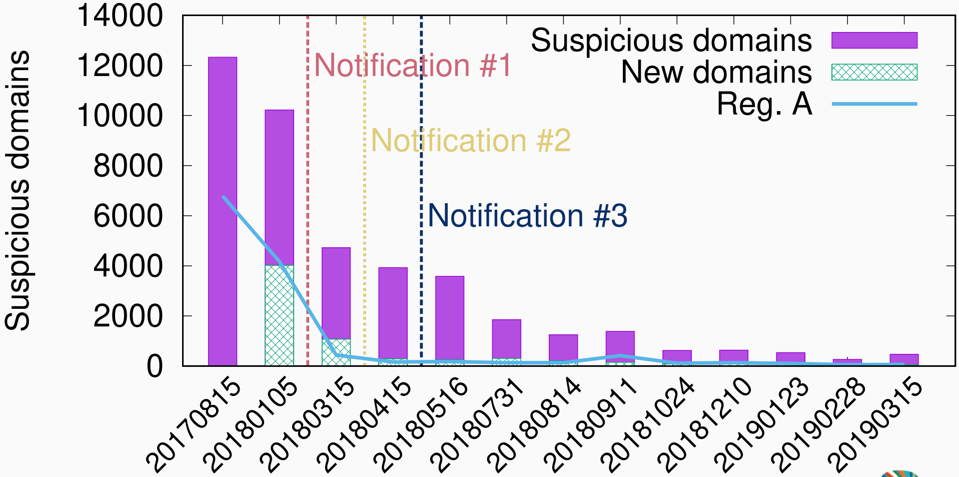
Our measurements

1. Get all `.nl` domain names (5.8M)
    - private data
2. Scrape their websites (if they have)
3. We deployed “state-of-the art” ML to detect them
    - simply count the number of brands on `<title>`

```
<title>Vans Schoenen On Sale 70% OFF |Geen verzendkosten</title>
```

- `if` > 5, then flag it
- (we precompiled a list of brands and discount words)

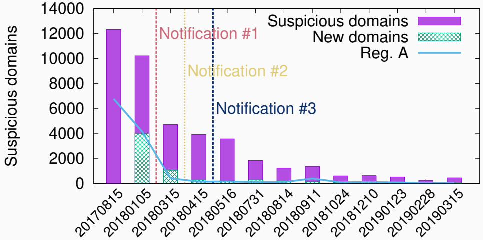
What did we find?



Takedown and lessons

- We could not take them down, legally
  - but registrars could, we notified them
- Win-win: applied academic skills to an industry problem
  - **Real-world impact**: Prevented people from getting scammed
  - Improved the zone security

More details: PAM2020 [2] paper



Outline

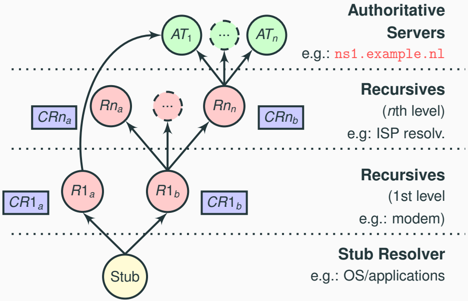
Case 1: Web Security

Case 2: Large Authoritative DNS servers Ops

Case 3: Time services on the Internet

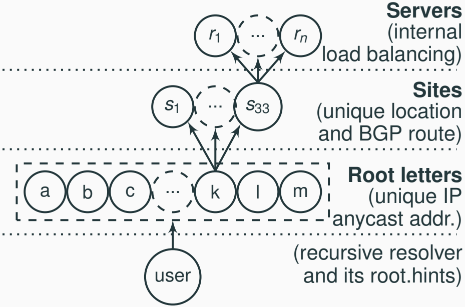
Industry + Academics working to solve OPs problems

- Two main types of DNS servers
- If **ALL** authoritative server fails, zone becomes unreachable
- Critical mission: it cannot fail
  - imagine `.nl` being down



Authoritative Servers Setup

- You can't have only one server
- The Root DNS servers have 13 addresses
    - but more than 700 physical machines and VMs
- Multiple layers of redundancy



Servers
(internal load balancing)
r_1
…
r_n
Sites
(unique location and BGP route)
s_1
…
s_{33}
Root letters
(unique IP anycast addr.)
a
b
c
…
k
l
m
user
(recursive resolver and its root.hints)

Figure 4: Root DNS structure, terminology, and replication levels.

Operators and Academics working together

- Industry/OPs:
  - SIDN `.nl` zone
  - B-ROOT `.` zone
- Academics:

SIDN LABS

TU Delft

USC Viterbi Information Sciences Institute

UNIVERSITY OF TWENTE.

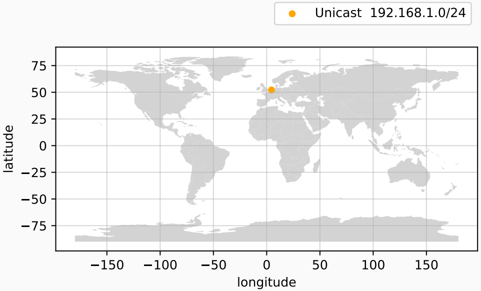
Research questions

1. How anycast reacts to DDoS? IMC2016
2. How resolvers choose Auth Servers? IMC2017
3. How cache protects DNS during DDoS? IMC2018
4. How to fine tune caching? IMC2019
5. Measure client latency from TPC traffic PAM2022

Summary for OPs folks: IETF RFC 9199

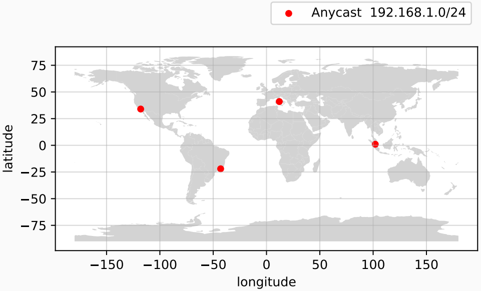
IP anycast

Unicast



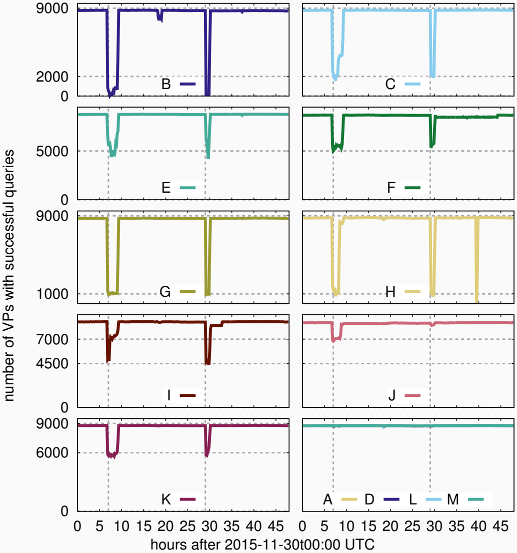
- One location
- All traffic to it

Anycast



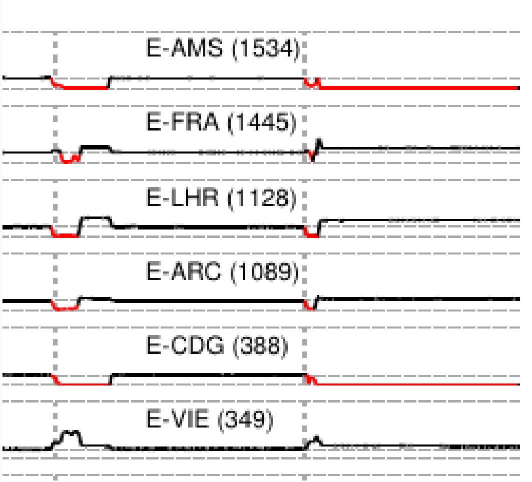
- Multiple locations
- Traffic distributed among them

Root DNS DDoS Nov 2015



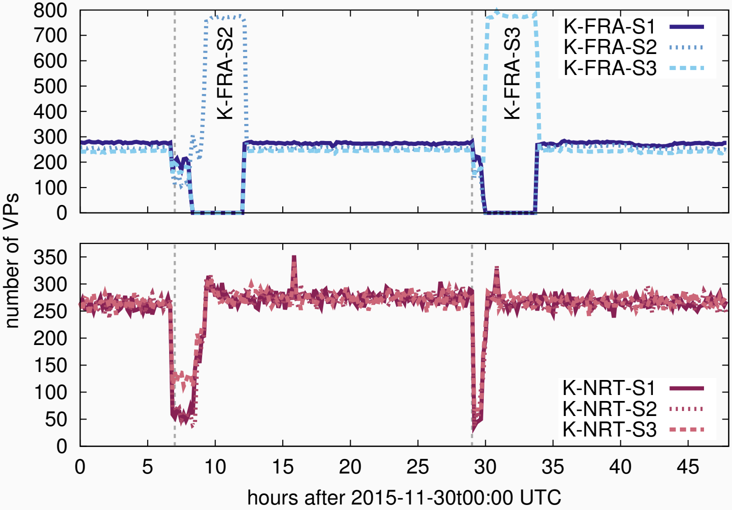
How anycast sites reacted?

Waterbed effect

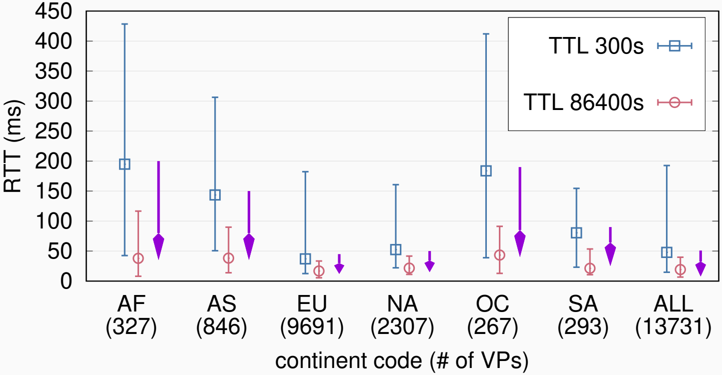


How individual servers reacted?

Two behaviors by load balancers



Caching and Latency: .uy latency reduced for all regions



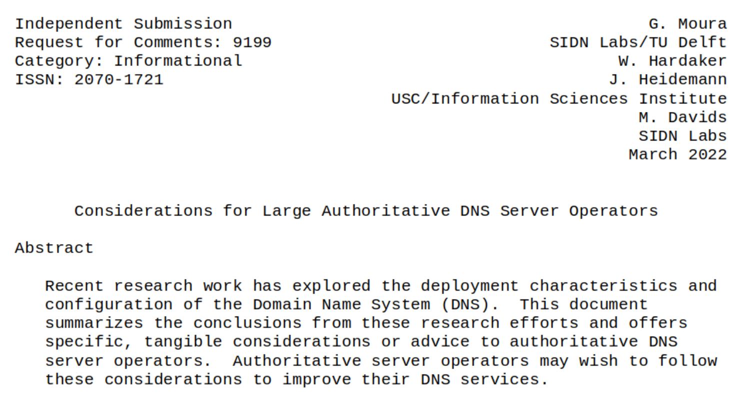
Longer TTL → longer caching → faster answers

Up to 150ms median latency reduction (AF)

See IMC2019

Summarizing recommendations to Operators

OPs folks prefer RFCs than papers

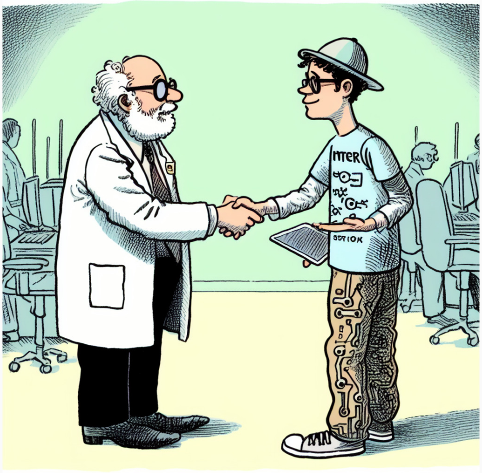


```
Independent Submission                                          G. Moura
Request for Comments: 9199                            SIDN Labs/TU Delft
Category: Informational                                      W. Hardaker
ISSN: 2070-1721                                             J. Heidemann
                                      USC/Information Sciences Institute
                                                               M. Davids
                                                               SIDN Labs
                                                              March 2022


      Considerations for Large Authoritative DNS Server Operators

Abstract

   Recent research work has explored the deployment characteristics and
   configuration of the Domain Name System (DNS).  This document
   summarizes the conclusions from these research efforts and offers
   specific, tangible considerations or advice to authoritative DNS
   server operators.  Authoritative server operators may wish to follow
   these considerations to improve their DNS services.
```

Takeway

- We solved DNS operation problems
- With provided operators with recommendations
- We helped to improve real world services
  - for real users

Academia + Industry working together



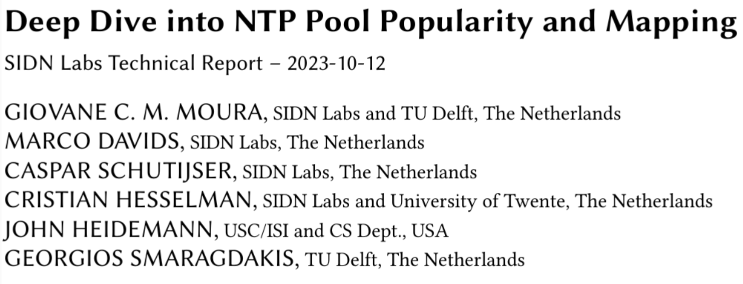
Outline

Case 1: Web Security

Case 2: Large Authoritative DNS servers Ops

Case 3: Time services on the Internet

Case 3: Time services on the Internet



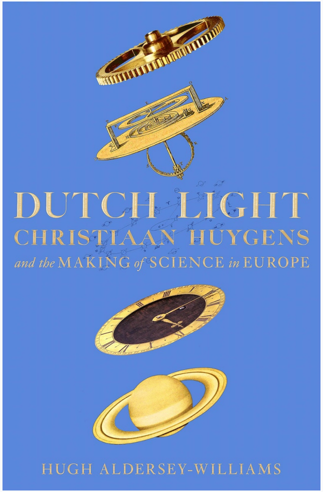
Deep Dive into NTP Pool Popularity and Mapping

SIDN Labs Technical Report – 2023-10-12

GIOVANE C. M. MOURA, SIDN Labs and TU Delft, The Netherlands
MARCO DAVIDS, SIDN Labs, The Netherlands
CASPAR SCHUTIJSER, SIDN Labs, The Netherlands
CRISTIAN HESSELMAN, SIDN Labs and University of Twente, The Netherlands
JOHN HEIDEMANN, USC/ISI and CS Dept., USA
GEORGIOS SMARAGDAKIS, TU Delft, The Netherlands

- Latest work (under review)
- First work since becoming Assistant Professor here
  - With Georgios, USC/ISI, SIDN Labs and Twente

Time in the Netherlands

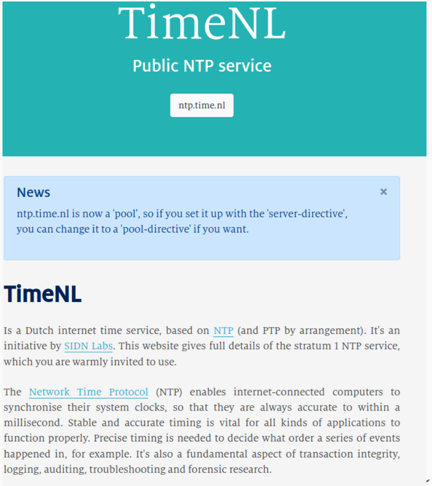


Time in the Netherlands: the pendulum clock

- Invented in 1656
- Gold standard until 1930!
- Lost its reign to quartz crystal oscillators and atomic clocks
- VSL (`https://vsl.nl`), have 4 atomic clocks
  - here on the campus (8 min by bike)
  - they provide the Netherlands Standard Time

Industry and academia

- SIDN operates `time.nl`, a tier-1 NTP service
- NTP is rather an overlooked research topic
- So we decided to look into it



TimeNL

Public NTP service

News

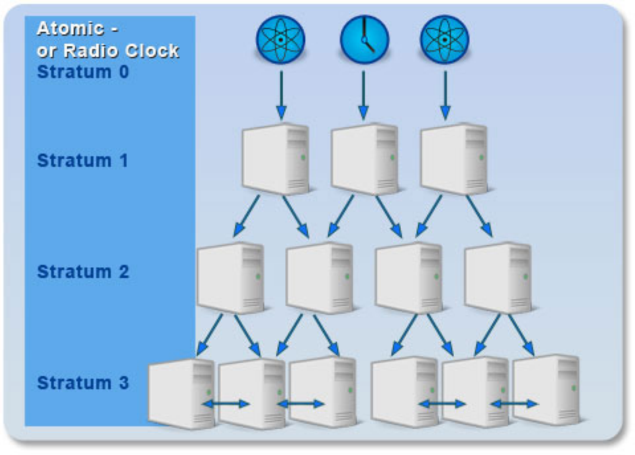
ntp.time.nl is now a 'pool', so if you set it up with the 'server-directive', you can change it to a 'pool-directive' if you want.

TimeNL

Is a Dutch internet time service, based on NTP (and PTP by arrangement). It's an initiative by SIDN Labs. This website gives full details of the stratum 1 NTP service, which you are warmly invited to use.

The Network Time Protocol (NTP) enables internet-connected computers to synchronise their system clocks, so that they are always accurate to within a millisecond. Stable and accurate timing is vital for all kinds of applications to function properly. Precise timing is needed to decide what order a series of events happened in, for example. It's also a fundamental aspect of transaction integrity, logging, auditing, troubleshooting and forensic research.

NTP synchronization



Time Sync on the Internet

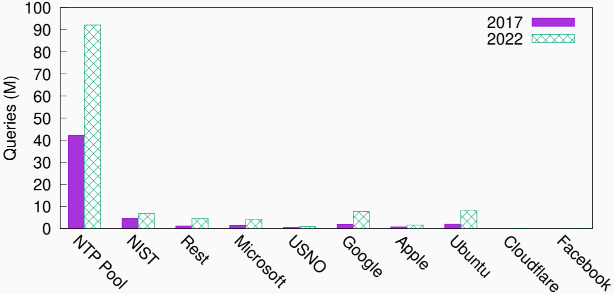
- NTP services
  - NIST
  - US Navy Naval Observatory
  - NTP Pool*
  - Later: Apple, Google, Cloudflare, Meta

NTP pool

- Volunteers who share their NTP servers
- 4000+ servers
- NTP Pool operators only run a DNS servers
  - volunteers run NTP servers
- Are they popular?

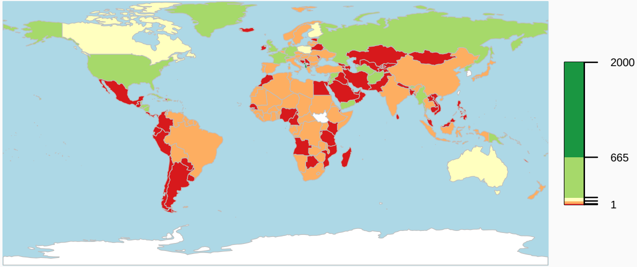
Measuring NTP Service Popularity Using DNS

- We analyzed Root DNS Server Traffic
- NTP pool tops all counts
    - Volunteers keeping the time on the Internet
- NTP pool is a bunch of volunteers



If NTP pool is king, how it map users to servers?

- The pool has 4000+ servers
- Which criteria it uses to map clients?

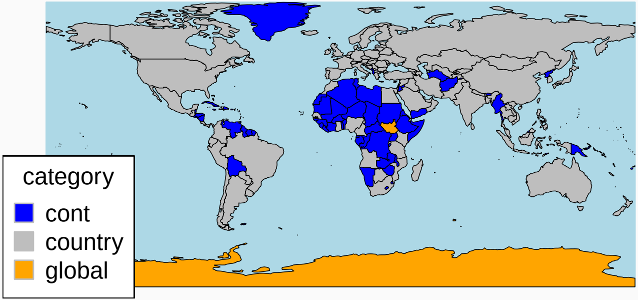


Number of NTP servers from NTP Pool clients will be served

- Very unfair mapping : red < 10 servers, orange < 50
- We suggested changes to GeoDNS, the DNS component of the Pool

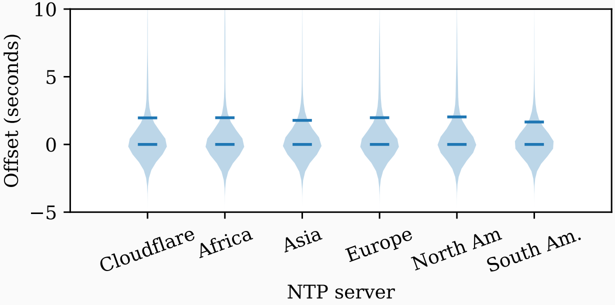
Easy to become a time provider a the entire country

- GeoDNS is to strict
- You can be these blue countries time keeper with one server

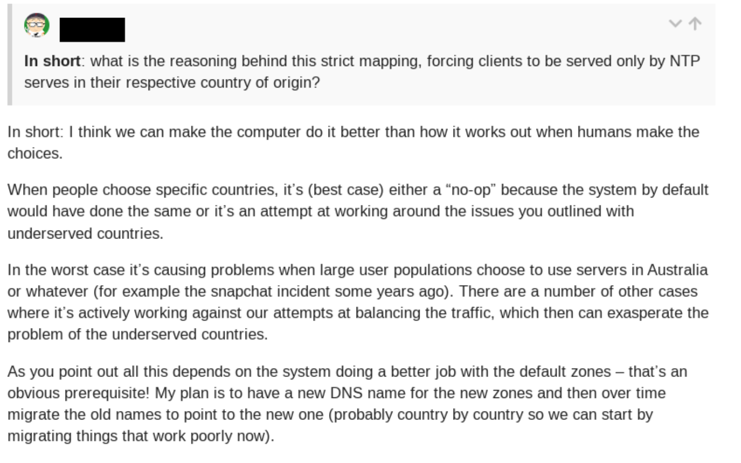


NTP Pool GeoDNS is based on a wrong assumption

- Far away servers cant' provide good service
- AND avoid asymmetric routing
- Most of Internet paths are asymmetric already
- We measured NTP offset from 132 VPs on countries only served by Cloudflare
    - same performance



We discussed with the operator



[redacted]

> **In short**: what is the reasoning behind this strict mapping, forcing clients to be served only by NTP serves in their respective country of origin?

In short: I think we can make the computer do it better than how it works out when humans make the choices.

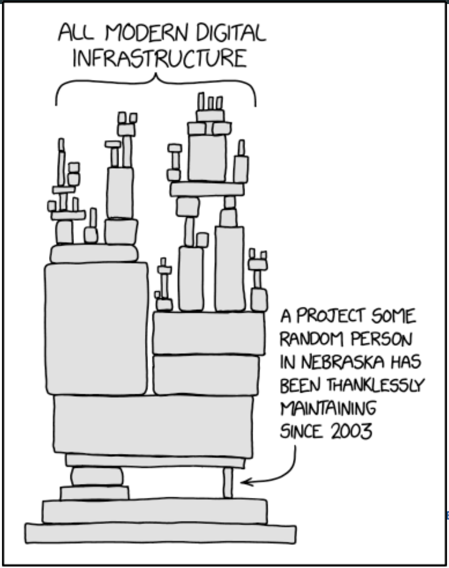
When people choose specific countries, it’s (best case) either a “no-op” because the system by default would have done the same or it’s an attempt at working around the issues you outlined with underserved countries.

In the worst case it’s causing problems when large user populations choose to use servers in Australia or whatever (for example the snapchat incident some years ago). There are a number of other cases where it’s actively working against our attempts at balancing the traffic, which then can exasperate the problem of the underserved countries.

As you point out all this depends on the system doing a better job with the default zones – that’s an obvious prerequisite! My plan is to have a new DNS name for the new zones and then over time migrate the old names to point to the new one (probably country by country so we can start by migrating things that work poorly now).

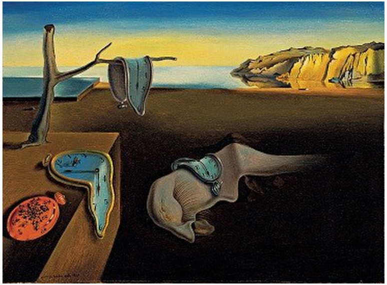
Takeway

- We looked into NTP and NTP pool
- We found several problems
- We discussed with operators
- They are working on a fix


ALL MODERN DIGITAL INFRASTRUCTURE
A PROJECT SOME RANDOM PERSON IN NEBRASKA HAS BEEN THANKLESSLY MAINTAINING SINCE 2003

Wrap-up

- Industry and Academia can do great works together
  - They have complementary skills
- We showed the cases where we did it:
  - Web security
  - DNS operations
  - NTP services and OPs
- Joint-work SIDN Labs and TU Delft
- Contact:
  - giovane-moura.nl
  - giovane.moura@sidn.nl, @tudelft.nl



The persistence of memory – Salvador Dali

References i

[1] Schmidle, N.

Inside the Knockoff-Tennis-Shoe Factory - The New York Times.

http://www.nytimes.com/2010/08/22/magazine/22fake-t.html, 2010.

[2] Wabeke, T., Moura, G. C. M., Franken, N., and Hesselman, C.

Counterfighting Counterfeit: detecting and taking down fraudulent webshops at a ccTLD.

In *Proceedings of the Passive and Active Measurement Workshop* (Eugene, OR, USA, 2020).

References ii

[3] Wall, D. S., and Large, J.

Jailhouse frocks: Locating the public interest in policing counterfeit luxury fashion goods.

The British Journal of Criminology 50, 6 (2010), 1094–1116 – http://ssrn.com/abstract=1649773.

[4] Wang, D. Y., Der, M., Karami, M., Saul, L., McCoy, D., Savage, S., and Voelker, G. M.

Search + seizure: The effectiveness of interventions on seo campaigns.

References iii

In *Proceedings of the 2014 Conference on Internet Measurement Conference* (New York, NY, USA, 2014), IMC '14, ACM, pp. 359–372.