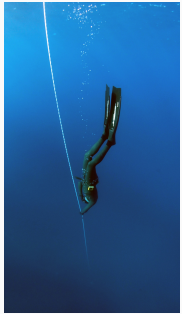
DNS: Unveiling the Critical Link in Internet Security and Exploring Diverse Use Cases

Giovane C. M.Moura
SIDN Labs and TU Delft

WTMC'23 – Opening Keynote
Delft, The Netherlands
2023-07-07

Today's Goals



No

img src: Unsplash

Yes

img src: wallpaperflare

Today's Goals

1. Show how can you use DNS in your research on:
   - Internet Security
   - Networking
2. Provide references and pointers
   - papers
   - datasets
   - (text in red is clicklable)

Yes

img src: wallpaperflare

Today's Goals

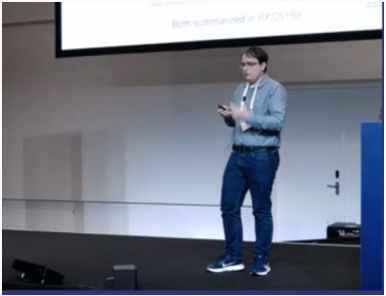
1. Show how can you use DNS in your research on:
   - Internet Security
   - Networking
2. Provide references and pointers
   - papers
   - datasets
   - (text in red is clicklable)

Yes

img src: wallpaperflare

$whoami

- Data Scientist at SIDN Labs
  - resarch team of SIDN, `.nl` registry
- Assistant Professor at TU Delft
  - my office at TU Delft is in this building :)
- Research focus on **operations**:
  - Internet Security
  - Networking
  - Systems
- PhD (2013, UTwente, NL)
- MSc (2008, UFRGS, BR)



Presentation @ RIPE86, Rotterdam, May 2023

Today's presentation

Counterfeit webshops

Logo Misusue

E-gov DNS

Wrap-up

Common reactions when people hear “DNS”

Reaction #1

Reaction #2

My hope for the day

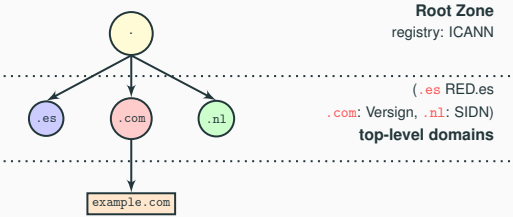
img src: Unsplash

(Slides will be online, content in red is clicklable link)

What is DNS?

- several protocols
- distributed database
- client-server-server architecture
- routing
- governance
- security
- performance
- 2000+ pages of documentation (DNS Camel)

DNS as a distributed database



- Each node in the tree is managed by a different organization
- Why?

A DNS registry and .nl

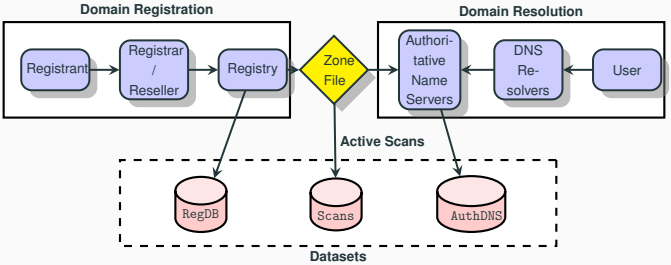


Figure 1: TLD operations: registration (left), domain resolution (right), and datasets.

Outline

Counterfeit webshops

Logo Misusue

E-gov DNS

Wrap-up

Back in 2016 ... strange websites

- 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?

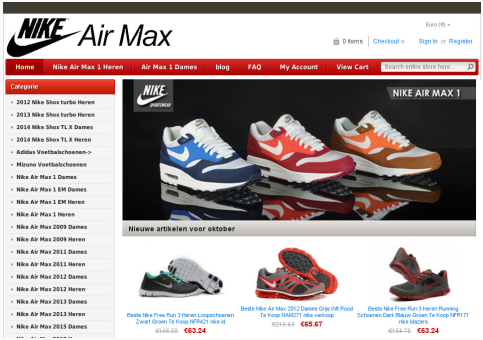


Figure 2: Screenshot of 2016 `.nl` website

Does anyone even buy from them?

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

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*?

- Counterfeit (fake) industry is **huge**: books, computers, shoes, bags
  - EU borders seizures 2016: 670 miliion EUR
  - US 2017: US$ 1.2 Billion
- Luxury goods have a massive demand

If you buy a fake from the street, you know it

- but not online
- so we got involved

What is their *business model*?

- The business model goes like this:
  1. Consumer demand [4]
  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?

- Back to 2016: we stumbled on them
- 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 [5]

How many were on the .nl zone?

- Back to 2016: we stumbled on them
- 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 [5]

How many were on the .nl zone?

- Back to 2016: we stumbled on them
- 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 [5]

Our measurements

1. Get all `.nl` domain names (5.8M)
   - private data
2. Scrape their websites (if they have)
   - We used DMap [6], we are trying to open it
3. We deployed "state-of-the art" ML to detect
   - 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)

Our measurements

1. Get all `.nl` domain names (5.8M)
   - private data
2. Scrape their websites (if they have)
   - We used DMap [6], we are trying to open it
3. We deployed “state-of-the art” ML to detect
   - 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)

Our measurements

1. Get all `.nl` domain names (5.8M)
   - private data
2. Scrape their websites (if they have)
   - We used DMap [6], we are trying to open it
3. We deployed “state-of-the art” ML to detect
   - 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)

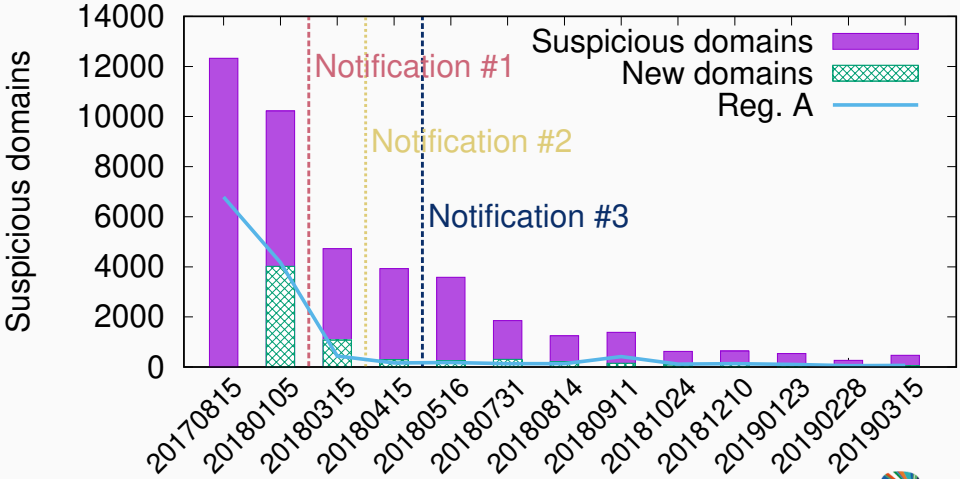
Our measurements

1. Get all `.nl` domain names (5.8M)
   - private data
2. Scrape their websites (if they have)
   - We used DMap [6], we are trying to open it
3. We deployed “state-of-the art” ML to detect
   - simply count the number of brands on `<title>`

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

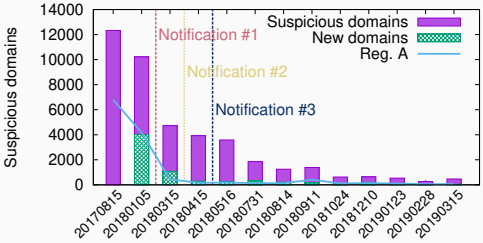
   - `if > 5`, then flag it
   - (we precompiled a list of brands and discount words)

What did we find?



How to take them down it?

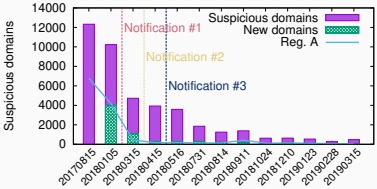
- We could not take them down
- But there was a way to validate them:
  1. Notify a registrar that registered the domain
  2. Ask them to verify the ID of the registrant
  3. If it fails, then they can suspend the domain



Lessons

1. How come does this even work?
   - This is to show they suffered little pressure
2. Why so many of these webshops?
   - it's unlikely there are that many counterfeiters
   - *Domains are cheap and disposable*
   - automation heavily used
   - 10 down does not even make a difference
3. Why 6K were registered with only one registrar?
   - API for automatic registration & good price

Take downs were effective, in partnership with our registrars

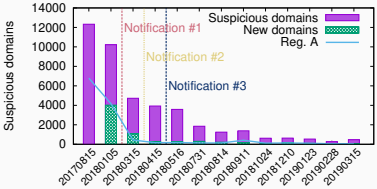


- Later they changed strategy, we had a new system
- See PAM2020 [3]

Lessons

1. How come does this even work?
   - This is to show they suffered little pressure
2. Why so many of these webshops?
   - it's unlikely there are that many counterfeiters
   - *Domains are cheap and disposable*
   - automation heavily used
   - 10 down does not even make a difference
3. Why 6K were registered with only one registrar?
   - API for automatic registration & good price

Take downs were effective, in partnership with our registrars

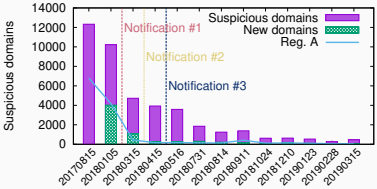


- Later they changed strategy, we had a new system
- See PAM2020 [3]

Lessons

1. How come does this even work?
   - This is to show they suffered little pressure
2. Why so many of these webshops?
   - it's unlikely there are that many counterfeiters
   - *Domains are cheap and disposable*
   - automation heavily used
   - 10 down does not even make a difference
3. Why 6K were registered with only one registrar?
   - API for automatic registration & good price

Take downs were effective, in partnership with our registrars

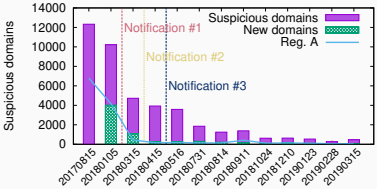


- Later they changed strategy, we had a new system
- See PAM2020 [3]

Lessons

1. How come does this even work?
   - This is to show they suffered little pressure
2. Why so many of these webshops?
   - it's unlikely there are that many counterfeiters
   - *Domains are cheap and disposable*
   - automation heavily used
   - 10 down does not even make a difference
3. Why 6K were registered with only one registrar?
   - API for automatic registration & good price

Take downs were effective, in partnership with our registrars

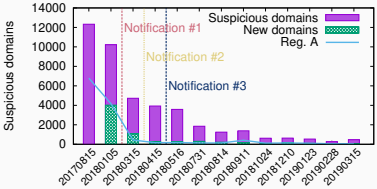


- Later they changed strategy, we had a new system
- See PAM2020 [3]

Lessons

1. How come does this even work?
   - This is to show they suffered little pressure
2. Why so many of these webshops?
   - it's unlikely there are that many counterfeiters
   - *Domains are cheap and disposable*
   - automation heavily used
   - 10 down does not even make a difference
3. Why 6K were registered with only one registrar?
   - API for automatic registration & good price

Take downs were effective, in partnership with our registrars

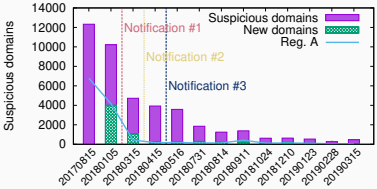


- Later they changed strategy, we had a new system
- See PAM2020 [3]

Lessons

1. How come does this even work?
   - This is to show they suffered little pressure
2. Why so many of these webshops?
   - it's unlikely there are that many counterfeiters
   - *Domains are cheap and disposable*
   - automation heavily used
   - 10 down does not even make a difference
3. Why 6K were registered with only one registrar?
   - API for automatic registration & good price

Take downs were effective, in partnership with our registrars

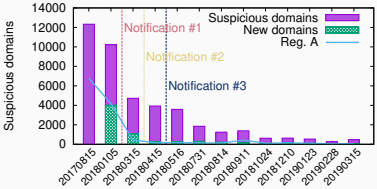


- Later they changed strategy, we had a new system
- See PAM2020 [3]

Lessons

1. How come does this even work?
   - This is to show they suffered little pressure
2. Why so many of these webshops?
   - it's unlikely there are that many counterfeiters
   - *Domains are cheap and disposable*
   - automation heavily used
   - 10 down does not even make a difference
3. Why 6K were registered with only one registrar?
   - API for automatic registration & good price

Take downs were effective, in partnership with our registrars



- Later they changed strategy, we had a new system
- See PAM2020 [3]

Outline

Counterfeit webshops

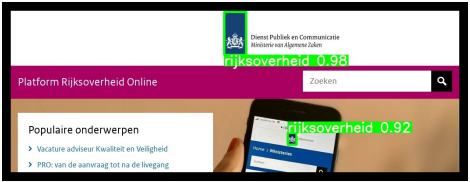
Logo Misusue

E-gov DNS

Wrap-up

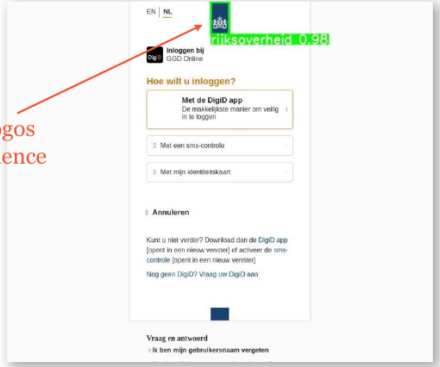
From text to logo detection: LogoMotive

- My colleagues did a study evaluating misuse of Dutch government logo
- It became a brand protection service
- See PAM2022 [2] paper

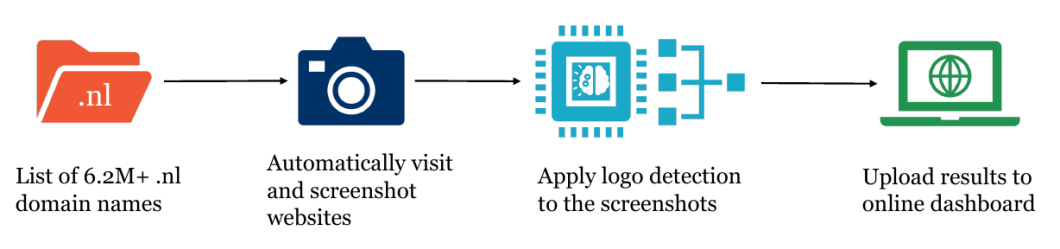


Detecting logos misuse with ML

Detected logos with confidence



How does LogoMotive work?



Generating training datasets

- We've used Yolo for image recognition
- It requires labeled data
- So we've generated it

	Value
Screenshots generated	64,893
Synthetic training samples	100,000
training set	95,000
validation set	5,000

Table 1: Datasets used for raining and validation.

Generating training datasets

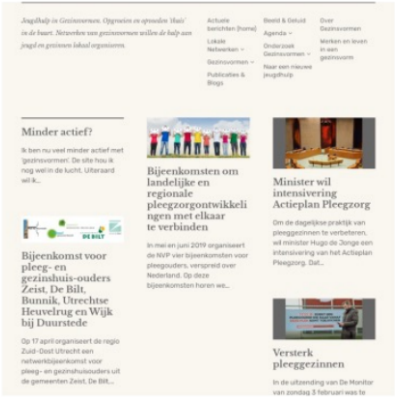
- We've used Yolo for image recognition
- It requires labeled data
- So we've generated it

	Value
Screenshots generated	64,893
Synthetic training samples	100,000
training set	95,000
validation set	5,000

Table 1: Datasets used for raining and validation.

Generating training datasets

+



=

Random screenshot **Resulting datapoint**

Evaluating the model

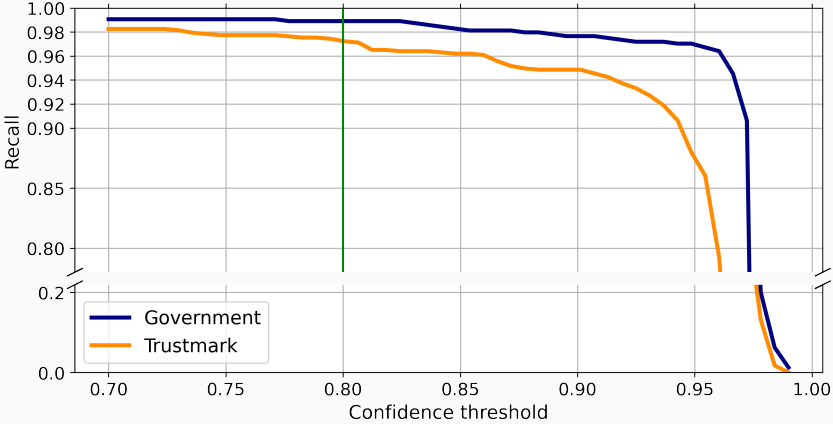


Figure 4: Recall performance of LogoMotive at confidence thresholds. The vertical line denotes our chosen threshold.

Results

Label	Full-Zone	Newly-Registered
Total	12862 (100.00%)	53
Without gov. logo (FP)	1164 (9.05%)	0 (0.00%)
With gov. logo (TP)	11698 (90.95%)	53 (100.0%)
Benign	10595 (82.37%)	32 (60.38%)
Government impersonation	151 (1.17%)	17 (32.09%)
Phishing	3 (0.02%)	3 (5.66%)
Potential threat	73 (0.57%)	9 (16.98%)
Other (false endorsements, satire, etc.)	75 (0.58%)	5 (9.43%)
Government domains	952 (7.40%)	4 (7.55%)
In portfolio	636 (4.94%)	2 (0.00%)
Not in portfolio	316 (2.46%)	2 (3.77%)
Added	109 (0.85%)	1 (1.89%)
Pending	207 (1.61%)	1 (1.89%)

Table 2: Manual validation results for government impersonation case study.

On the paper

- See PAM2022 [2] paper for more details
- There was a second case study
- It became a brand protection service

You can also DIY!

You don't need private data:

1. Get DNS zone files
   - Sweden's `.se` is open
   - ICANN CZDS has all gTLDs, and `.com`, `.net`, and `.org`
   - Ask your country ccTLD
2. Get an open-source crawler
   - Mercator from DNSBelgium
3. Figure out problems
   - Detect X impersonation

Outline

Counterfeit webshops

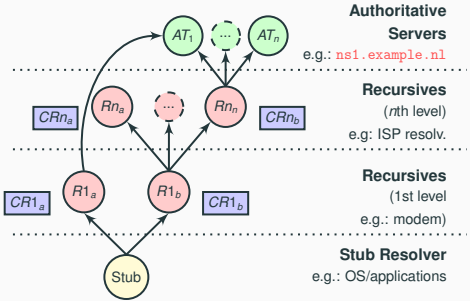
Logo Misusue

E-gov DNS

Wrap-up

DNS Servers and DNS infrastructure

- Two main types of DNS servers
- If authoritative server fails, zone becomes unreachable
- (previous exampled covered contents, this is infrastructure)



E-gov

- Governments increasingly use Internet for communication with citizens (e-gov)
- E-gov provide crucial services

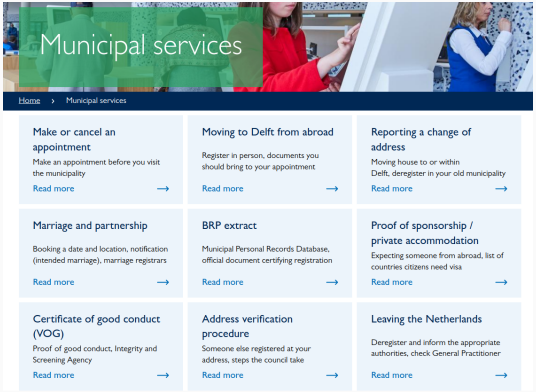
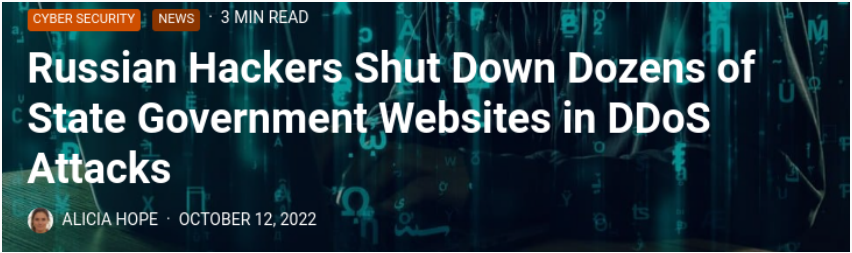


Figure 5: Delft (local government) residents e-gov

When e-gov breaks



source: CPO Magazine

"Russian hackers took responsibility for a wave of cyber attacks that knocked dozens of state government websites offline.
Several states, including Colorado, Connecticut, Kentucky, and Mississippi, were impacted by the politically-motivated cyber attacks ..."

E-gov is fully dependent on DNS

- E-gov provide crucial services
- Internet as core communications fabric of modern societies.
- E-gov is fully dependent on DNS

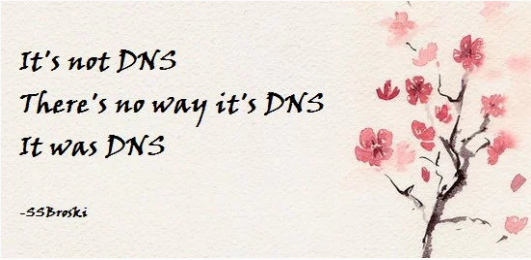


Figure 6: A haiku about DNS.
Source: Cyberciti

DNS Engineering for resilience

- DNS has been designed for resilience
  - multiple layers of redundancy
- Deploying those features is not easy/cheap
- Configuration errors may go unnoticed
  - system will still work
  - until it breaks

Source: Unsplash

Research Question

Are e-gov DNS serves configured following best-practices for robustness?

Approach: Internet measurements

Research Question

Are e-gov DNS serves configured following best-practices for robustness?

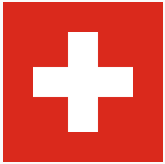
Approach: Internet measurements

Our contribution

1. E-gov DNS infrastructure evaluation for four countries
    - using active measurements
2. A comparative analysis among them
3. Recommendations for improvement

The Netherlands

Switzerland



Sweden

United States

Datasets

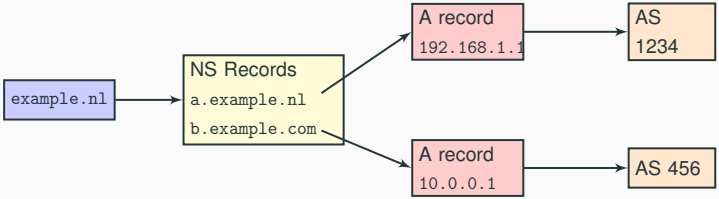
Country	Netherlands .nl	Sweden .se	Switzerland .ch	United States .gov
e-gov domains (SLD)	602	614	3971	7972
Population	17.4M	10.4M	8.7M	332.9M

Results: single points of failure (SPoF)

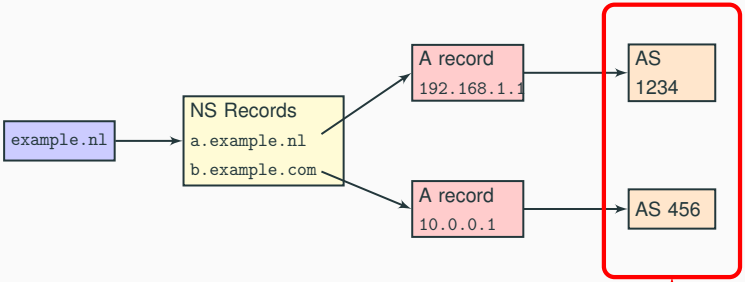
- Don´t put all your eggs in one basket
  - We will look into diff basket types

Source: Unsplash

First SPOF: single DNS providers



First SPOF: single DNS providers



Two ASes ∼ 2 DNS providers

First SPOF: single DNS providers

	Netherlands	Sweden	Switzerland	United States
second-level domains	602	614	3971	7972
Responsive	601	609	3546	7911
single provider(v4/v6)	43% /55%	41%/41%	43%/54%	82%/ 55%

- **US: $\sim$ 80% single DNS provider**

“But this is a bogus metric!”

- “I’ll put everything in the **cloud**”
- But even clouds occasionally fail:
  - Dyn 2016
  - AWS Route 53 - 2019
- Even Amazon.com does not use AWS for DNS:

```
pdns1.ultradns.net.
ns4.p31.dynect.net.
ns2.p31.dynect.net.
pdns6.ultradns.co.uk.
ns1.p31.dynect.net.
ns3.p31.dynect.net.
```

“But this is a bogus metric!”

- “I’ll put everything in the **cloud**”
- But even clouds occasionally fail:
  - Dyn 2016
  - AWS Route 53 - 2019
- Even Amazon.com does not use AWS for DNS:

```
pdns1.ultradns.net.
ns4.p31.dynect.net.
ns2.p31.dynect.net.
pdns6.ultradns.co.uk.
ns1.p31.dynect.net.
ns3.p31.dynect.net.
```

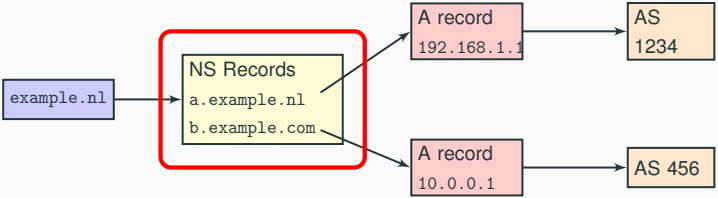
DNS centralization: who are these DNS providers

Netherlands		Sweden		Switzerland		United States	
ASN	e-gov	ASN	e-gov	ASN	e-gov	ASN	e-gov
Transip	112	Loopia	47	Infomaniak	278	GoDaddy	1215
CLDIN	39	Tele2	23	Swisscomm	115	Cloudflare	909
QSP	28	Microsoft	21	Novatrend	100	Amazon	676
Solvinity	8	Telia	21	Abraxas	97	Akamai	334
SSC-ICT	8	Telia	19	Metanet	91	Tiggee	316

Table 3: Top 5 DNS providers for e-gov domains

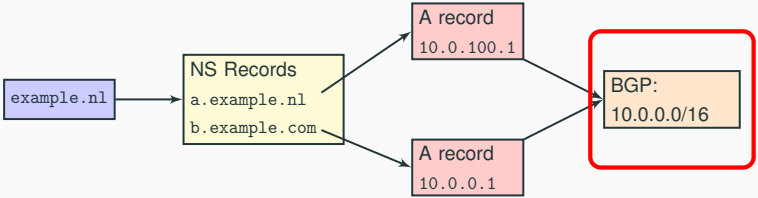
Most DNS providers are **local**

Second SPoF: single DNS server



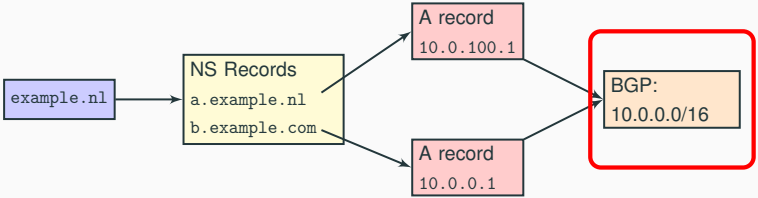
- RFC1034 (35 years old!) mandates at least two NS records
- We found 6 `.gov` domains that did have a single NS record
- We notified the `.gov` registry, 3 fixed it (2023-05-09)

Third SPoF: BGP prefixes



- If two DNS servers share the same prefix, they are not topologically diverse
  - they share the same infrastructure
- We map the IP addresses of each NS to their prefixes

Third SPoF: BGP prefixes

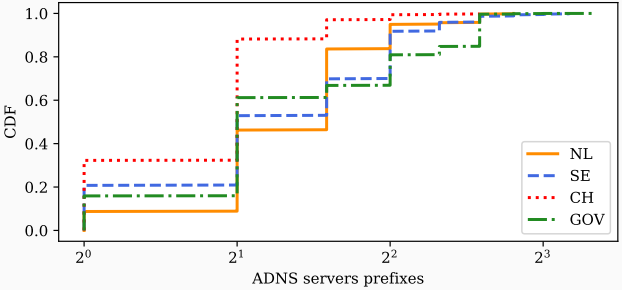


One BGP prefix = same location

- If two DNS servers share the same prefix, they are not topologically diverse
  - they share the same infrastructure
- We map the IP addresses of each NS to their prefixes

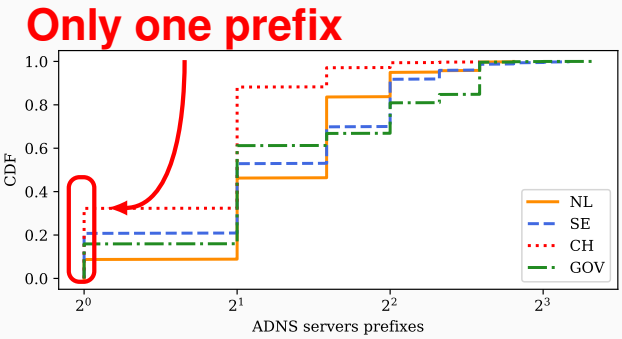
Third SPoF: BGP prefixes

- Switzerland: 1/3 e-gov domains have a single prefix
- NL, SE, US: $< 20\%$

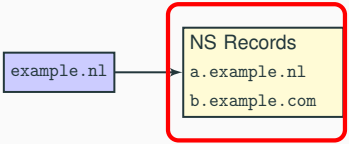


Third SPoF: BGP prefixes

- Switzerland: 1/3 e-gov domains have a single prefix
- NL, SE, US: $< 20\%$

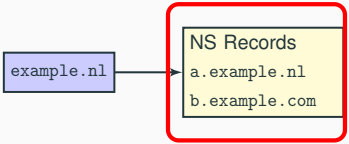


Fourth SPoF: Number of TLDs



- NS records depend on top-level domains (TLDs)
- Having more than one TLD protect you fail TLD failures
  - Warning: it's TLDs for NS records, not the domains themselves

Fourth SPoF: Number of TLDs

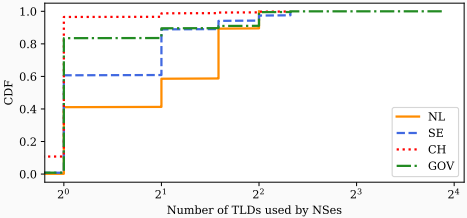


Two TLDs: `.nl` and `.com`

- NS records depend on top-level domains (TLDs)
- Having more than one TLD protect you fail TLD failures
  - Warning: it's TLDs for NS records, not the domains themselves

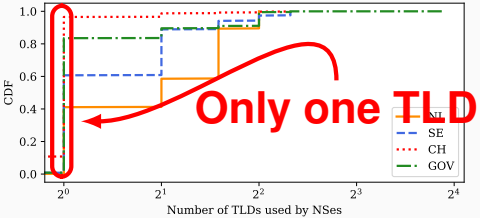
Fourth SPoF: Number of TLDs

- Switzerland e-gov mostly uses only one TLD
- Netherlands is the most diverse
- All four countries can diversity still



Fourth SPoF: Number of TLDs

- Switzerland e-gov mostly uses only one TLD
- Netherlands is the most diverse
- All four countries can diversity still



TLD dependency

	Netherlands	Sweden	Switzerland	United States
1	170 (.nl)	483 (.se)	609 (.ch)	2507 (.com)
2	69 (.net)	100 (.net)	190 (.com)	1541 (.net)
3	26 (.com)	82 (.com)	150 (.net)	894 (.gov)
4	12 (.eu)	14 (.info)	19 (.org)	485 (.org)
5	4 (.be)	8 (.org)	12 (.de)	302 (.us)

Table 4: Most used TLD by e-gov ADNS severs.

- Most use their own TLD, then .com and .net

Extra features that improve resilience (RFC9199)

1.IP Anycast

- Covered in Moura16b

2.DNS Time-to-live (TTLs)

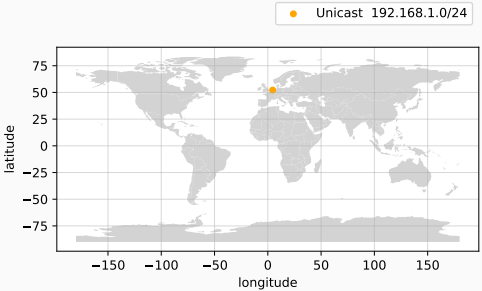
- covered in Moura18b, Moura19b

```
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.
```

Both summarized in RFC9199

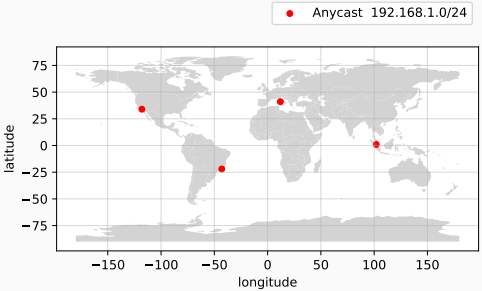
IP anycast

Unicast



- One location
- All traffic to it

Anycast

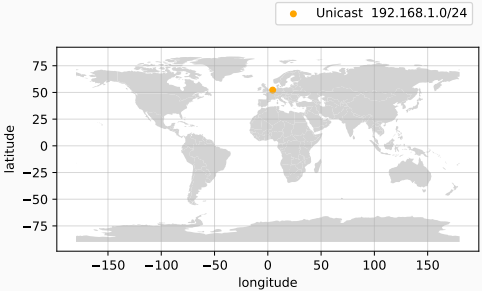


- Multiple locations
- Traffic distributed among them

Anycast is more resilient to DDoS (Moura16b)

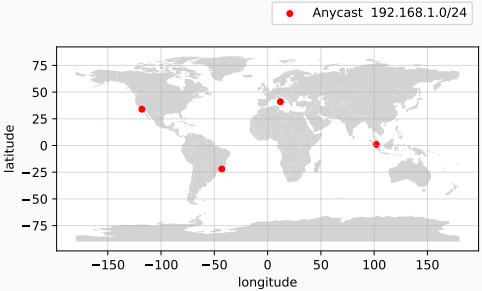
IP anycast

Unicast



- One location
- All traffic to it

Anycast

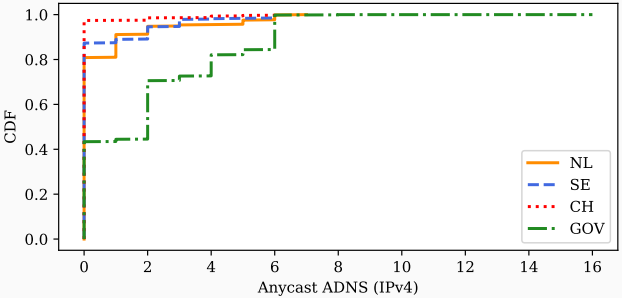


- Multiple locations
- Traffic distributed among them

Anycast is more resilient to DDoS (Moura16b)

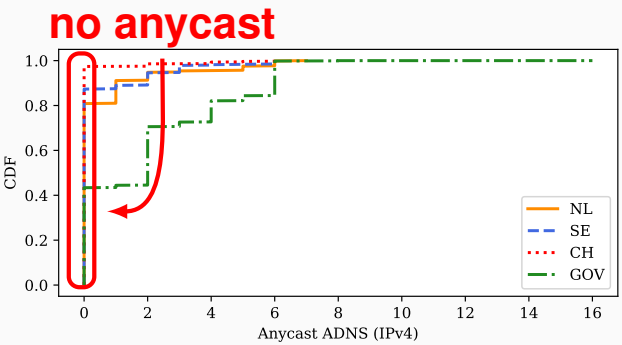
IP anycast adoption on e-gov

- Good: 58% US `.gov` domains have anycast
- Not so good: very few Swiss e-gov domains have anycast
- Sweden and the Netherlands have around 20% of anycast servers



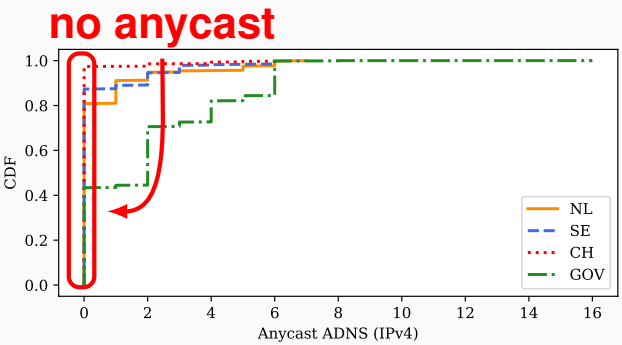
IP anycast adoption on e-gov

- Good: 58% US `.gov` domains have anycast
- Not so good: very few Swiss e-gov domains have anycast
- Sweden and the Netherlands have around 20% of anycast servers



IP anycast adoption on e-gov

- Good: 58% US `.gov` domains have anycast
- Not so good: very few Swiss e-gov domains have anycast
- Sweden and the Netherlands have around 20% of anycast servers



DNS time-to-live (TTL)

- TTLs control how long DNS records should stay in resolver's cache
- Last resort when everything else fails (Moura18b)
- Current recommendations: use at least a couple of hours TTL

Source: Unsplash

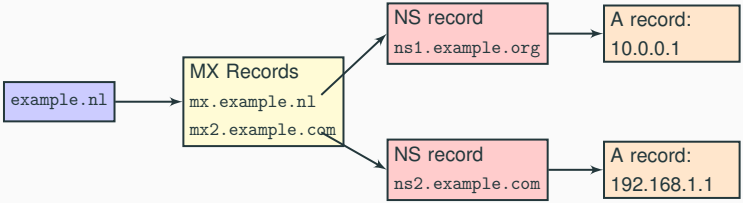
DNS time-to-live (TTL)

	Netherlands	Sweden	Switzerland	United States
NS TTL				
Median	10800	3600	3600	10800
A/AAAA TTL				
Median	3047	3600	3600	28800

E-gov e-mail DNS

- So far we've looked into E-gov DNS for **web**
- E-mail is also an important e-gov sevice
- Now we turn to measure the resilience of e-gov DNS for e-mail

E-gov e-mail DNS



- For e-mail we first retrieve their MX records, and proceed as previous

E-gov e-mail DNS

Country	Netherlands .nl	Sweden .se	Switzerland .ch	United States .gov
e-gov domains (SLD)	602	614	3971	7972
Outlook	164 (39%)	205 (37%)	425 (22.1%)	2243 (41%)

- E-gov E-mail uses mosly Microsoft regardless of the country
- Why? Maybe they seek for more traditional solutions
  - more in the paper[PDF]

Recommendations for e-gov DNS

- **Diversify:** more DNS providers, more NS records, more prefixes, different TLDs for NS records
- **Deploy** anycast for more robust services
- **Reconsider** low TTL values

Robust (1900 years old) infrastructure in Segovia, Spain. Src: Wikipedia

Lessons

- Many e-gov domains are not following the recommendation for robust services
- This creates unnecessary risk
- We hope our findings prompt the responsible operators to improve the redundancy and resilience of e-gov DNS

Robust (1900 years old) infrastructure in Rome, Italy. Src: Wikipedia

Full paper: Sommese22a

Outline

Counterfeit webshops

Logo Misusue

E-gov DNS

Wrap-up

Wrap-up

- DNS offers great opportunities for research
  - both in contents and infrastructure
- We've covered examples that can be easily reproduced
- I hope these examples motivate folks
- Contact:
  - giovane-moura.nl
  - giovane.moura@sidn.nl, @tudelft.nl

Slides:

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] VAN DEN HOUT, T., WABEKE, T., MOURA, G. C. M., AND HESSELMAN, C.

Logomotive: detecting logos on websites to identify online scams - a tld case study.

In *Passive and Active Measurement* (2022).

References ii

[3] 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).

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

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

References iii

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

[5] 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.

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

References iv

[6] Wullink, M., Moura, G. C., and Hesselman, C.

Dmap: Automating domain name ecosystem measurements and applications.

In *Proceedings of the IEEE Network Traffic Monitoring and Analysis Conference* (Vienna, Austria, June 2018), IEEE, pp. 1–8.