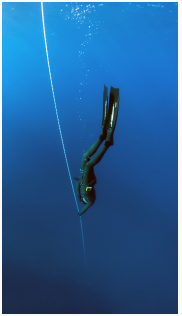
Internet Security in Practice

Perspectives from a DNS Operator

Giovane C. M.Moura
SIDN Labs and TU Delft

Bachelor CSE – CSE3220 – Guest Lecture
Delft, The Netherlands
2024-04-04

Today's Goals



No

img src: Unsplash

Yes

img src: wallpaperflare

Today's Goals

1. **Security = Economics**
   - two case studies
2. **Security = People**
   - one case study

From an **operator's** perspective

(But what is an operator?)

$whoami

- Data Scientist at SIDN Labs
  - research team of SIDN, `.nl` registry
- Assistant Professor at TU Delft
- Research focus on **operations**

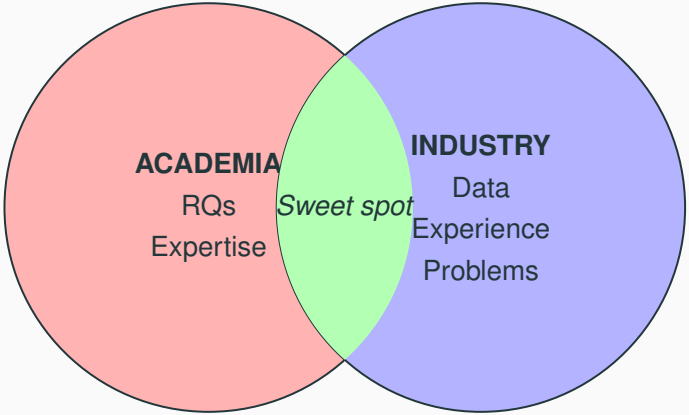
(Slides will be online, content in red is a clickable link)

Stereotypes

Academics seen by industry

Industry seen by academics

In reality



- Google has a nice paper on Industry/Academia collaboration

Example of Industry/Academia collaboration

- We (SIDN Labs) teamed up with SIDN OPs and B-Root OPs (USC/ISI)
- Goal: solve many open questions in DNS operations
- Outcome: 7 papers, and RFC9199

Independent Submission
Request for Comments: 9199
Category: Informational
ISSN: 2070-1721

G. Moura
SIDN Labs/TU Delft
W. Hardaker
J. Heidemann
USC/Information Sciences Institute
M. Davids
SIDN Labs

March 2022

Considerations for Large Authoritative DNS Server Operators

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

Today's presentation

Working on a DNS operator

Security = Economics

- Case 1: Counterfeit webshops
- Case 2: Online (logo) impersonation

Security = People

- Case 3: Vulnerability Disclosure

Outline

Working on a DNS operator

Security = Economics

- Case 1: Counterfeit webshops
- Case 2: Online (logo) impersonation

Security = People

- Case 3: Vulnerability Disclosure

Common reactions when people hear “DNS”

Reaction #1

Common reactions when people hear “DNS”

Reaction #2

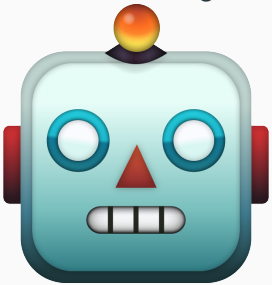
Working for an operator taught me a lot

1. They know WAY more than any academic
2. They focus is to run their systems
3. They appreciate research contributions
4. Their feedback is better than any reviewer in papers I ever had

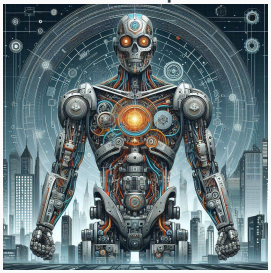
DNS and college

CSE gets like what? 30min about DNS?

DNS in college:



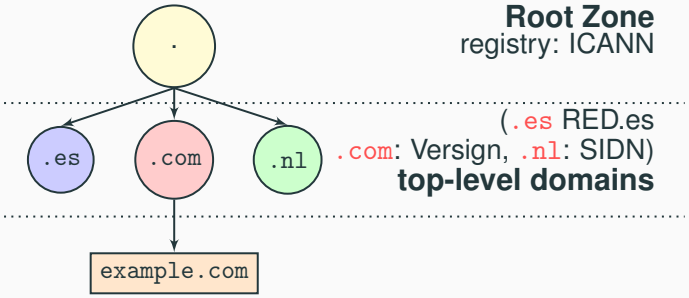
DNS in an operator:



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

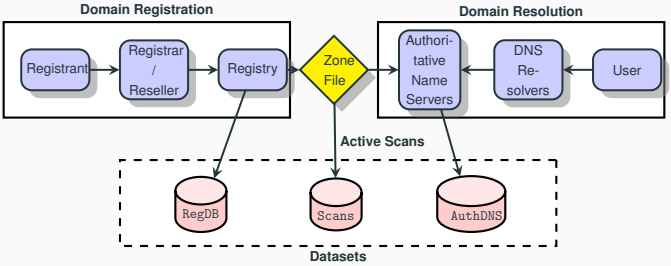


Figure 1: TLD operations: registration (left), domain resolution (right), and datasets.

Outline

Working on a DNS operator

Security = Economics

- Case 1: Counterfeit webshops
- Case 2: Online (logo) impersonation

Security = People

- Case 3: Vulnerability Disclosure

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?

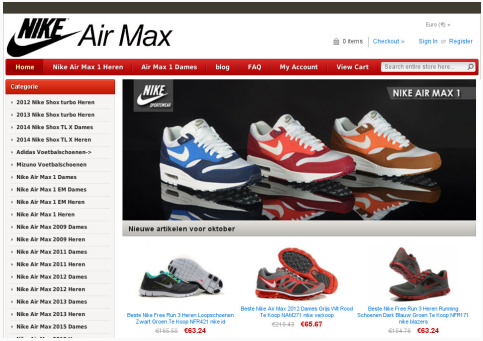


Figure 2: Screenshot of 2016 `.nl` website

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 the 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 the 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 the 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
- Border seizures:
  - EU 2022: € 2B
  - US 2023: $ 1.5B
- 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 [7]
  2. Manufacturing in China [3]
  3. These webshops connect both of them

Security is about Economics

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 [8]

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 [8]

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 [8]

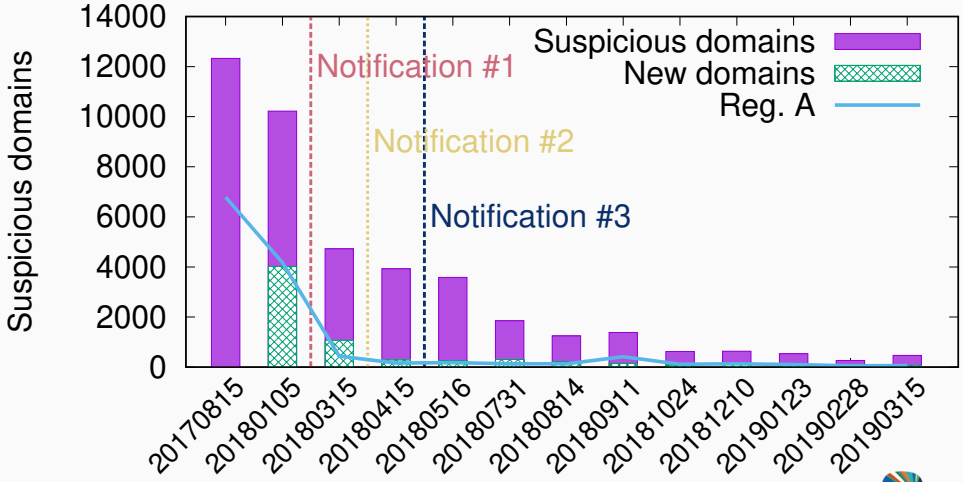
Our measurements

1. Get all `.nl` domain names (5.8M)
   - private data
2. Scrape their websites (if they have)
   - We used DMap [9], 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>
```

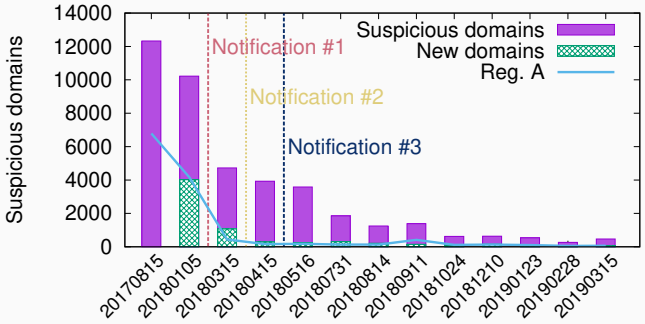
   - `if > 5`, then flag it
   - (we precompiled a list of brands and discount words)

What did we find?



They were taken down

- We could not take them down
- We notified registrars; they could
- **Scams removed from the .nl zone.**



More info: See PAM2020 paper [6]

Outline

Working on a DNS operator

Security = Economics

- Case 1: Counterfeit webshops
- Case 2: Online (logo) impersonation

Security = People

- Case 3: Vulnerability Disclosure

Online Impersonation

- Many websites display numerous logos.
- These logos often imply endorsements.
  - More logos, more money?
  - Security = Economics
- However, anyone can place any logo on their website.
- This raises the question: **Are logos being misused?**

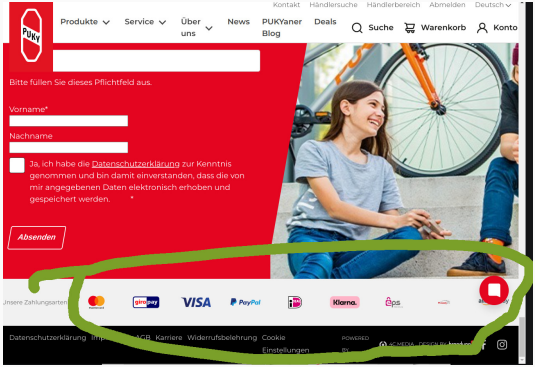


Figure 4: Legit website with Logos

Detecting Logos != Detecting text

You need:

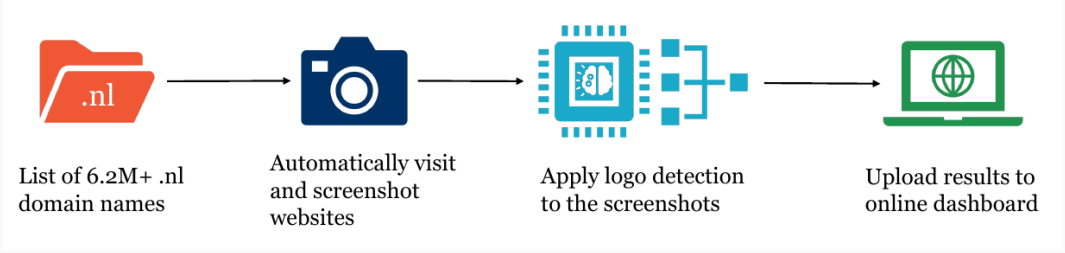
1. A list of websites
2. Visit them and “find” the logos:
    - Download each element OR
    - Generate a screenshot
3. Detect the logo *somehow*
    - You can use ML for that

Case study: Government of Netherlands Logo misuse

- My colleagues at SIDN Labs teamed up with a agency of the NL Government
- Goal: identify misuse of Rijksoverheid logo
- See PAM2022 [5] paper
- We focused on the `.nl` zone, which SIDN runs

Logo Rijksoverheid

How does LogoMotive work?



Detecting logos misuse with YoLo

- YoLo gives you a confidence score
  - from 0 to 1
- You choose your threshold based on FPs

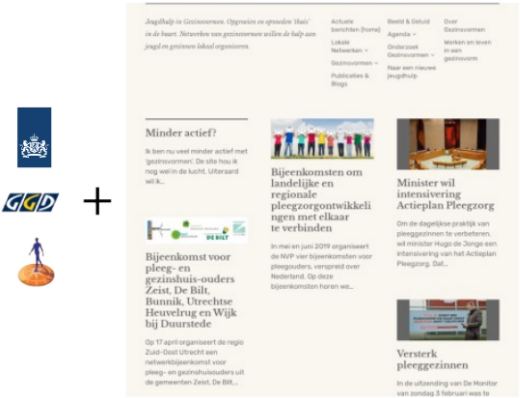
Generating training datasets

- YoLo 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

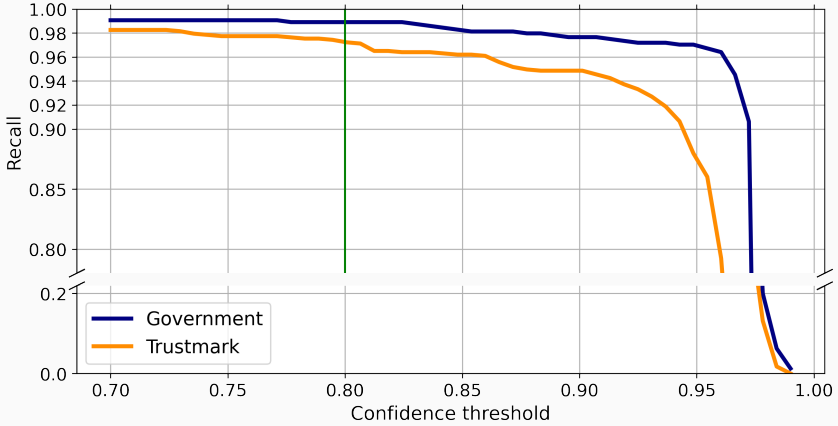


Figure 5: Recall performance of LogoMotive at confidence thresholds. The vertical line denotes our chosen threshold.

Results (thanks folks who validated it)

Total	**12862 (100.00%)**
Without gov. logo (FP)	1164 (9.05%)
With gov. logo (TP)	11698 (90.95%)
Benign	10595 (82.37%)
Government impersonation	151 (1.17%)
Phishing	3 (0.02%)
Potential threat	73 (0.57%)
Other (false endorsements, satire, etc.)	75 (0.58%)
Government domains	952 (7.40%)
In portfolio	636 (4.94%)
Not in portfolio	316 (2.46%)
Added	109 (0.85%)
Pending	207 (1.61%)

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

On the paper

- See PAM2022 [5] paper for more details
- There was a second case study

LogoMotive 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

Working on a DNS operator

Security = Economics

- Case 1: Counterfeit webshops
- Case 2: Online (logo) impersonation

Security = People

- Case 3: Vulnerability Disclosure

Vulnerability Disclosure

- “In practice, the theory is different”

(my Electricity and Magnetism professor at college)

Vulnerability Disclosure: Theory

Vulnerability Disclosure: Practice

La persistència de la memòria, Salvador Dali, 1931

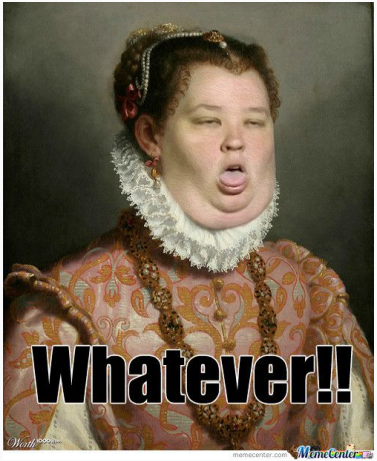
There is not even terminology consensus

What's the different between:

1. Private Disclosure
2. Public Disclosure
3. Full Disclosure
4. Responsible Disclosure
5. Coordinated Vulnerability Disclosure

Private Disclosure

- You tell only the vendor
- They decide to do whatever they want
- Commonly used in the past
- Outcomes:
  - Being ignored
  - Legal threats [4]



Full Disclosure

- The opposite of private disclosure:
  - You tell everyone, everything
- Only way to bring public scrutiny to vulnerabilities[4]
- It removes the veil of secrecy

In between both: “Responsible” or Coordinated Disclosure

- This is simply Full disclosure with an embargo:
- You give a vendor some time to fix it
  - US Cert suggests: 45 days
  - Google Project Zero: 90 days
- After it, you are “free” do disclose it
  - But WOULD you?
  - Imagine you vs one of the Big Tech companies?

Evolving terminology: use Coordinated, phase out Responsible

- “Responsible” disclosure implies a moral duty on whoever found the bug
- The responsible is the vendor! They created the bug
- Coordinated Vulnerability Disclosure (CVD) is the preferred term
- It removes the onus on the researcher and has not moralistic label

CVD is used by the NCSC-NL

Software bugs: there's plenty

- 0.5 to 25 bugs per 1000 LoC [1]
- CVE catalogs vulnerabilities
- No clear end in sight
  - software becomes more complex
  - weak incentives to make secure software

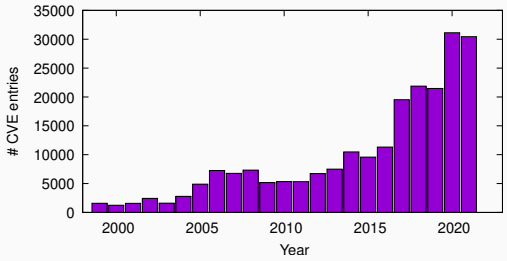


Figure 6: Yearly vulnerabilities listed by CVE.

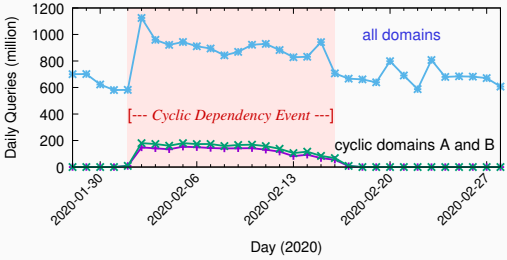
What do if you find a software bug?

What do if you find a software bug?

- Keep it private
- Sell it (See HackerOne)
  - This one will not fix the issue
  - Can empower attackers elsewhere
  - 1M USD for 0-day IOS bugs
- Disclose it
  - the most ethical choice

So we found one bug

- It affected Google Public DNS
- It caused 50% traffic increase on New Zealand's `.nz` DNS server
- Important:
  - Third-parties were the victims, not GDNS
- What to do?
  - There were not many papers telling 1st hand experience
  - Uncharted territory



TsuNAME Vulnerability

- Clients or resolvers would loop
- It could overwhelm authoritative DNS servers
- Google has far more capacity than most operators
- An attacker could aim GDNS at DNS servers

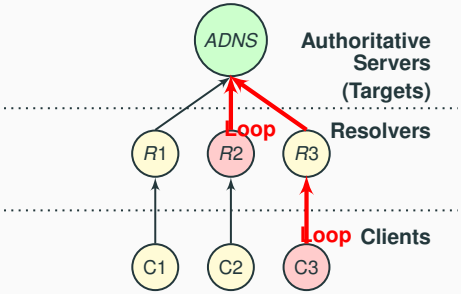
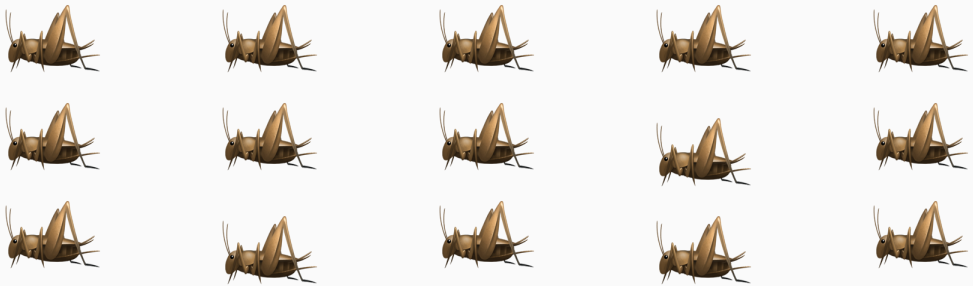


Figure 7: TsuNAME attack.

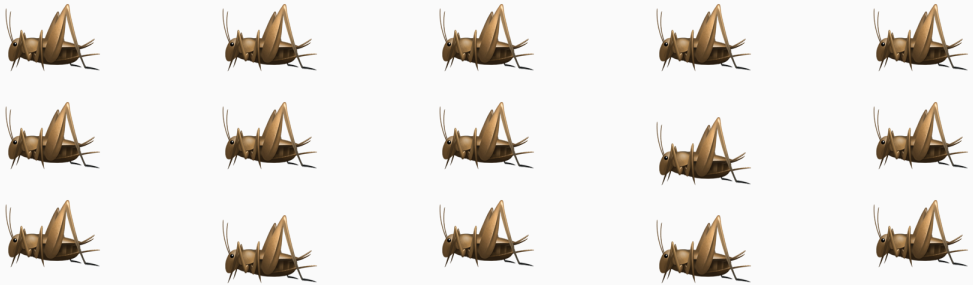
So what did we do?

- We knew some folks at GDNS
- So we notified them personally
- (Private disclosure)

So what happened?



So what happened?



But why?

Security = People

We made lots of mistakes

- So we wrote a paper about it [2]
- 1st hand experience
- And lessons learned

TsuNAME disclosure timeline

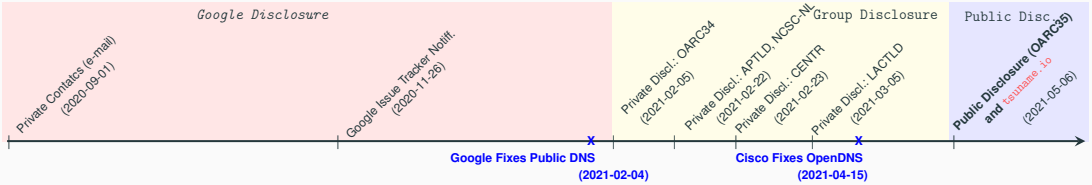


Figure 8: TsuNAME disclosure timeline

Lesson 1: Full disclosure improves security for everyone

- `.nz` has 50% traffic increase on TsuNAME
- We wonder why there had been no public reports on it
  - given it had a big damage potential
- We decide to disclose it
- It was ultimately fixed
- Improved security for everyone

Lesson 2: Disclosure has ethical implications

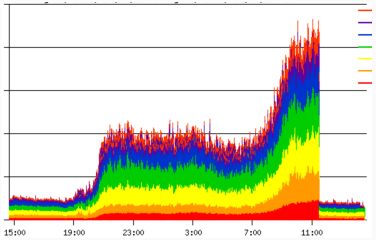
- Don’t disclose and others can become victims
- Our initial private disclosure only to Google did not work
  - it was our mistake
- We first notified all vendors in the group disclosure phase
  - they all release reports on the vulnerability

Lesson 3: ask help to reduce burden

- We disclosed it in many venues, in 4 languages
- It took lots of time and energy
- US-CERT has a vulnerability disclosure coordinator to help
  - so you don’t have to do it yourself
  - they take away all emotional/legal burden

Lesson 4: You don’t have the complete picture

- During our Q&A at OARC34 group disclosure, two ccTLD operators told us they had been victims of it before
- .nz had 50% traffic increase, an european ccTLD had 1000%.
- The other said it had tried private disclosure many times
  - we could not verify it
  - but is an example of why private disclosure does not work



Lesson 5: Prepare for stressful reactions

You can’t make everyone happy

- Positive reactions: Google, BIND, Cisco OpenDNS, Unbound
- Negative: one operator said it was fear-mongering, other said it was a known problem
  - it was partially known, but not at this scale
  - there’s a IETF draft now that covers it
- The primary goal is not to please everyone but to fix the problem

Improving the disclosure process

1. Clarify vendor roles and timeframes:
    - Most guidelines don't cover roles
    - Vendors can sit on a disclosure
    - Their bug system is also vague timeline wise:
        - TsuNAME on Google: "P2 issues need to be addressed on a reasonable timescale"
2. Update and endorse CVD guidelines
    - We need guidelines that protect individuals who disclosure
    - With clear timeframes
    - And concise

Security= People

Conclusions

- We covered two cases where Security = Economics
- One case where Security= People
- Operators have a lot to gain from academia and vice-versa
- It's a win-win situation for both
- More info: https://sidnlabs.nl
  - You can do an MSc Thesis internship with us
  - or work with us

References i

[1] Habib, A., and Pradel, M.

How many of all bugs do we find? a study of static bug detectors.

In *Proceedings of the 33rd ACM/IEEE International Conference on Automated Software Engineering* (New York, NY, USA, 2018), ASE '18, Association for Computing Machinery, p. 317–328.

[2] Moura, G. C. M., and Heidemann, J.

Vulnerability disclosure considered stressful.

SIGCOMM Comput. Commun. Rev. 53, 2 (jul 2023), 2–10.

References ii

[3] SCHMIDLE, N.

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

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

[4] SCHNEIER, B.

Schneier: Full Disclosure of Security Vulnerabilities a 'Damned Good Idea', 2007.

References iii

[5] 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).

[6] 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 iv

[7] 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.

[8] 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.

[9] 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.