DarkDNS: Revisiting the Value of Rapid Zone Update

RAFFAELE SOMMESE, GAUTAM AKIWATE,
ANTONIA AFFINITO, MORITZ MULLER,
MATTIJS JONKER, KC CLAFFY

AIMS-GMI3S – 26/06/2024

ICANN CZDS

- With the expansion of new generic TLDs (gTLDs), ICANN **mandated** zone files to be accessible through a simplified and centralized process: ICANN's **Centralized Zone Data Service**.
- Most of the gTLDs approved by ICANN!
- Facilitated a lot of research in DNS resilience, infrastructure and abuse.
- However,.....one snapshot a day?!

The problem of snapshots

Sunday, June 23, 2024, 6 PM

Monday, June 24, 2024, 6 PM

The problem of snapshots

Sunday, June 23, 2024, 8:54 PM

The need of more granular data

- Quickly detecting newly registered domains (e.g., for DNS Abuse detection)
- DNS Hijacking detection
- Live DNS infrastructural changes

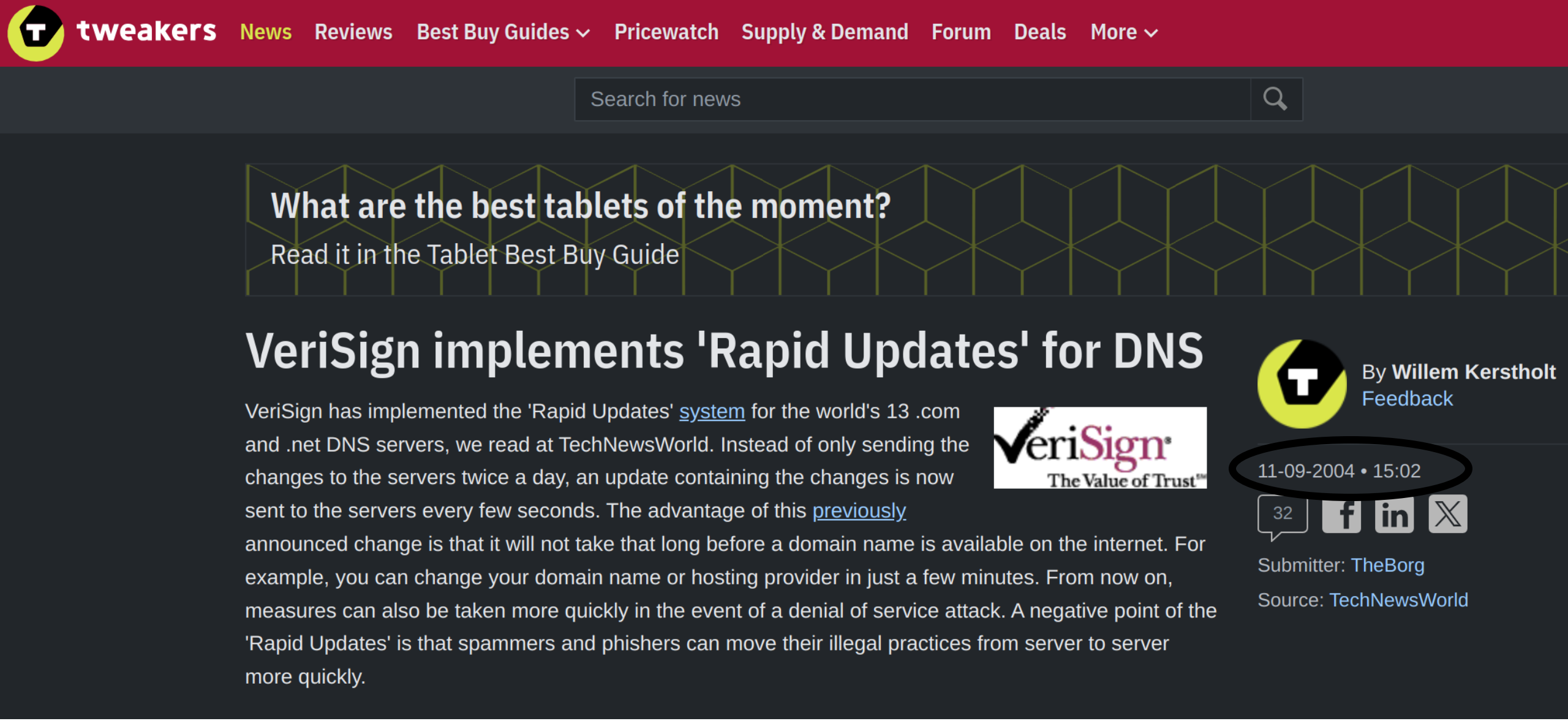
‘Short, Brutal Lives’: Life Expectancy for Malicious Domains

October 1, 2018 TH Author 0 Comments

Using a cooling-off period for domain names can help catch those registered by known bad actors.

Domain Name System (DNS) pioneer Paul Vixie for more than three years has been calling for a “cooling off” period for newly created Internet domain names as a way to deter cybercrime and other abuses. Domain names registered and spun up in less than a minute only encourage and breed malicious activity, he argues, and placing them in a holding pattern for a few minutes or hours can help vet them and catch any registered by known spammers and other bad actors.

Vixie — who is founder and CEO of threat intelligence firm Farsight Security — and his team have now taken an up-close look at the life cycle of new Internet domains, and their findings shine new light on the lifespan of malicious and suspicious domains. “Most of them die young, and most of them die after living short, brutal lives,” he says of newly created domains.



A blast from the past

7/13/2024

Alternatives?!

Passive DNS Data

- DomainTools Newly Observed Domain Names Feed

Limitations:

- Only "active queried" domains
- Commercial data source
- Limited availability

What else?


@old_master
That boy is our last hope...
No. There is another.

Computer Science > Networking and Internet Architecture

[Submitted on 4 Sep 2023]

This Is a Local Domain: On Amassing Country-Code Top-Level Domains from Public Data

Raffaele Sommese, Roland van Rijswijk-Deij, Mattijs Jonker

Domain lists are a key ingredient for representative censuses of the Web. Unfortunately, such censuses typically lack a view on domains under country-code top-level domains (ccTLDs). This introduces unwanted bias: many countries have a rich local Web that remains hidden if their ccTLDs are not considered. The reason ccTLDs are rarely considered is that gaining access -- if possible at all -- is often laborious. To tackle this, we ask: what can we learn about ccTLDs from public sources? We extract domain names under ccTLDs from 6 years of public data from Certificate Transparency logs and Common Crawl. We compare this against ground truth for 19 ccTLDs for which we have the full DNS zone. We find that public data covers 43%-80% of these ccTLDs, and that coverage grows over time. By also comparing port scan data we then show that these public sources reveal a significant part of the Web presence under a ccTLD. We conclude that in the absence of full access to ccTLDs, domain names learned from public sources can be a good proxy when performing Web censuses.

POSTER

Poster: Through the ccTLD Looking Glass: Mining CT Logs for Fun, Profit and Domain Names

Authors: Raffaele Sommese and Mattijs Jonker | Authors Info & Claims

IMC '23: Proceedings of the 2023 ACM on Internet Measurement Conference • October 2023 • Pages 714 - 715
https://doi.org/10.1145/3618257.3624994

Published: 24 October 2023 Publication History

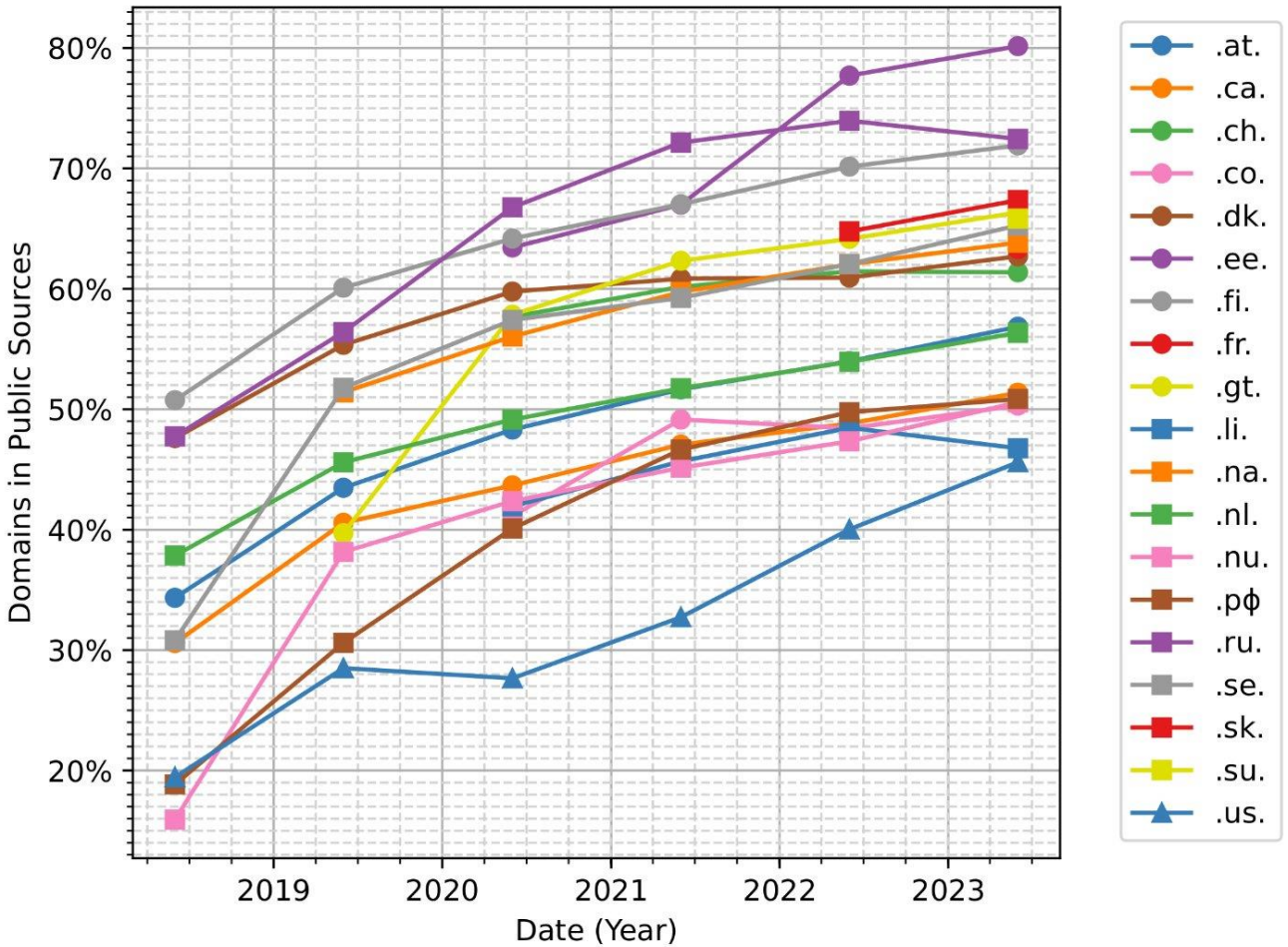
Computer Science > Networking and Internet Architecture

[Submitted on 4 Sep 2023]

This Is a Local Domain: On Amassing Country-Code Top-Level Domains from Public Data

Raffaele Sommese, Roland van Rijswijk-Deij, Mattijs Jonker

Domain lists are a key ingredient for representative censuses of the Web. Unfortunately, such censuses typically lack a view on domains under country-code top-level domains (ccTLDs). This introduces unwanted bias: many countries have a rich local Web that remains hidden if their ccTLDs are not considered. The reason ccTLDs are rarely considered is that gaining access -- if possible at all -- is often laborious. To tackle this, we ask: what can we learn about ccTLDs from public sources? We extract domain names under ccTLDs from 6 years of public data from Certificate Transparency logs and Common Crawl. We compare this against ground truth for 19 ccTLDs for which we have the full DNS zone. We find that public data covers 43%-80% of these ccTLDs, and that coverage grows over time. By also comparing port scan data we then show that these public sources reveal a significant part of the Web presence under a ccTLD. We conclude that in the absence of full access to ccTLDs, domain names learned from public sources can be a good proxy when performing Web censuses.



POSTER

Poster: Through the ccTLD Looking Glass: Mining CT Logs for Fun, Profit and Domain Names

Authors: Raffaele Sommese and Mattijs Jonker | Authors Info & Claims

IMC '23: Proceedings of the 2023 ACM on Internet Measurement Conference • October 2023 • Pages 714 - 715
https://doi.org/10.1145/3618257.3624994

Published: 24 October 2023 Publication History

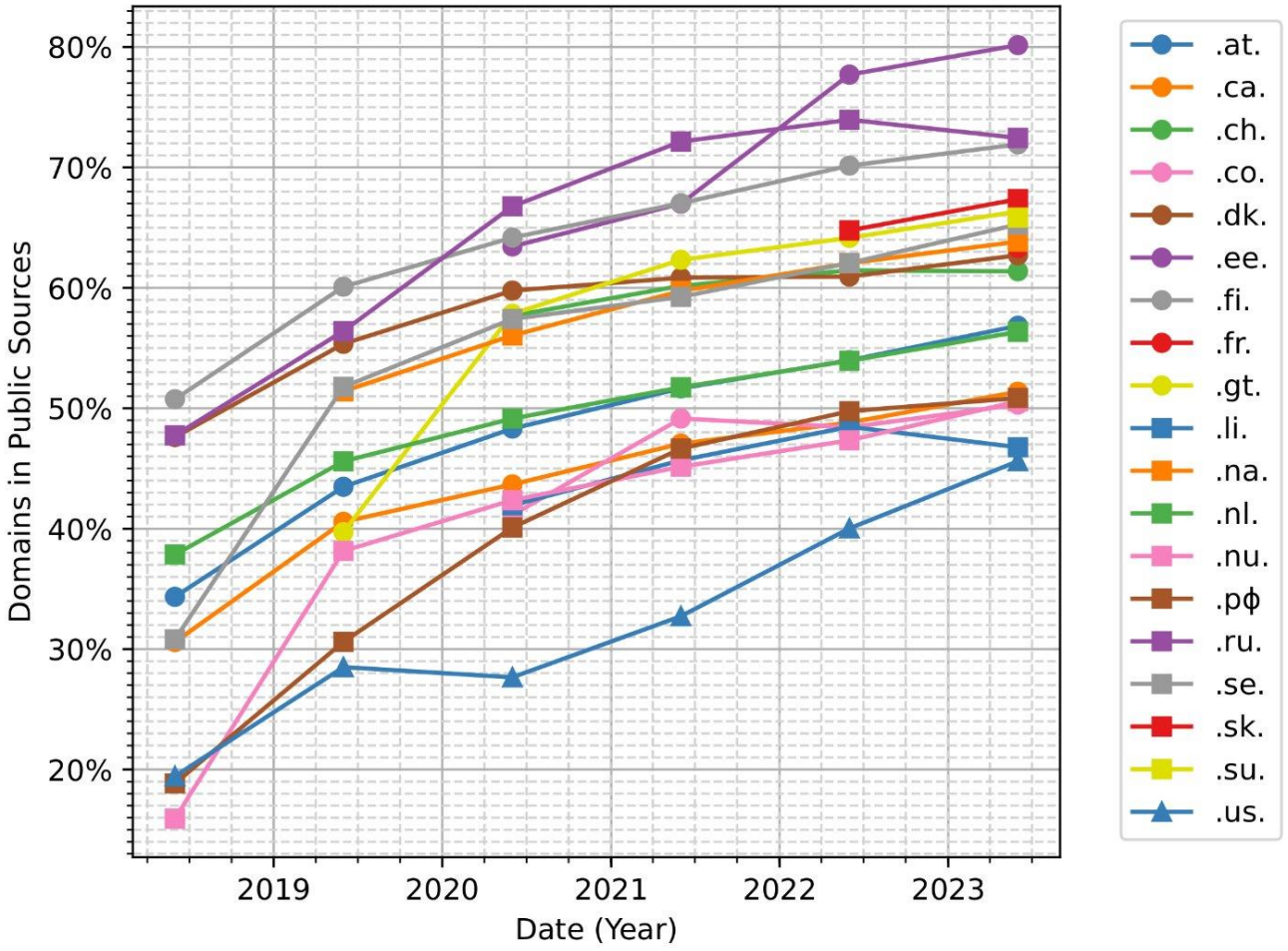
Computer Science > Networking and Internet Architecture

[Submitted on 4 Sep 2023]

This Is a Local Domain: On Amassing Country-Code Top-Level Domains from Public Data

Raffaele Sommese, Roland van Rijswijk-Deij, Mattijs Jonker

Domain lists are a key ingredient for representative censuses of the Web. Unfortunately, such censuses typically lack a view on domains under country-code top-level domains (ccTLDs). This introduces unwanted bias: many countries have a rich local Web that remains hidden if their ccTLDs are not considered. The reason ccTLDs are rarely considered is that gaining access -- if possible at all -- is often laborious. To tackle this, we ask: what can we learn about ccTLDs from public sources? We extract domain names under ccTLDs from 6 years of public data from Certificate Transparency logs and Common Crawl. We compare this against ground truth for 19 ccTLDs for which we have the full DNS zone. We find that public data covers 43%-80% of these ccTLDs, and that coverage grows over time. By also comparing port scan data we then show that these public sources reveal a significant part of the Web presence under a ccTLD. We conclude that in the absence of full access to ccTLDs, domain names learned from public sources can be a good proxy when performing Web censuses.



POSTER

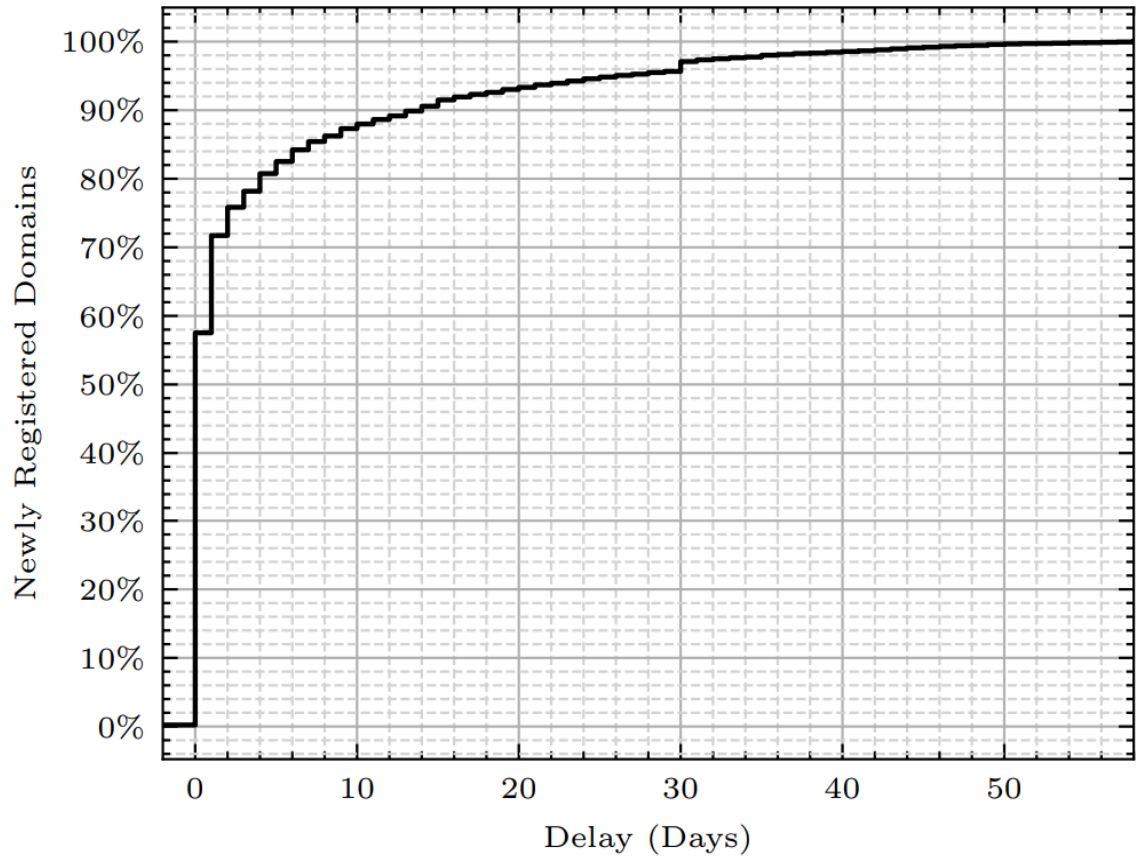
Poster: Through the ccTLD Looking Glass: Mining CT Logs for Fun, Profit and Domain Names

Authors: Raffaele Sommese and Mattijs Jonker | Authors Info & Claims

IMC '23: Proceedings of the 2023 ACM on Internet Measurement Conference • October 2023 • Pages 714 - 715
https://doi.org/10.1145/3618257.3624994

Published: 24 October 2023 Publication History

Check for updates



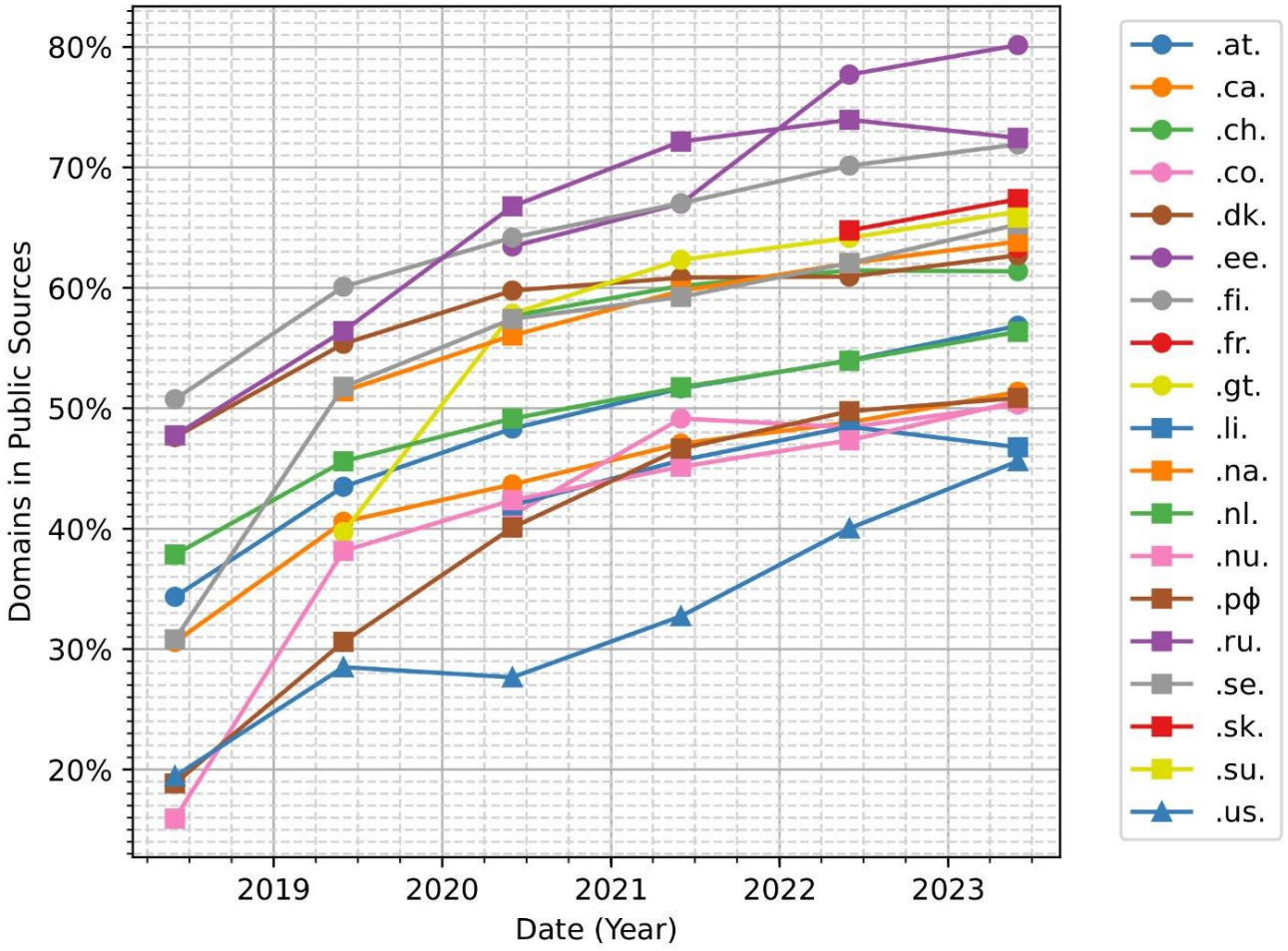
Computer Science > Networking and Internet Architecture

[Submitted on 4 Sep 2023]

This Is a Local Domain: On Amassing Country-Code Top-Level Domains from Public Data

Raffaele Sommese, Roland van Rijswijk-Deij, Mattijs Jonker

Domain lists are a key ingredient for representative censuses of the Web. Unfortunately, such censuses typically lack a view on domains under country-code top-level domains (ccTLDs). This introduces unwanted bias: many countries have a rich local Web that remains hidden if their ccTLDs are not considered. The reason ccTLDs are rarely considered is that gaining access -- if possible at all -- is often laborious. To tackle this, we ask: what can we learn about ccTLDs from public sources? We extract domain names under ccTLDs from 6 years of public data from Certificate Transparency logs and Common Crawl. We compare this against ground truth for 19 ccTLDs for which we have the full DNS zone. We find that public data covers 43%-80% of these ccTLDs, and that coverage grows over time. By also comparing port scan data we then show that these public sources reveal a significant part of the Web presence under a ccTLD. We conclude that in the absence of full access to ccTLDs, domain names learned from public sources can be a good proxy when performing Web censuses.



POSTER

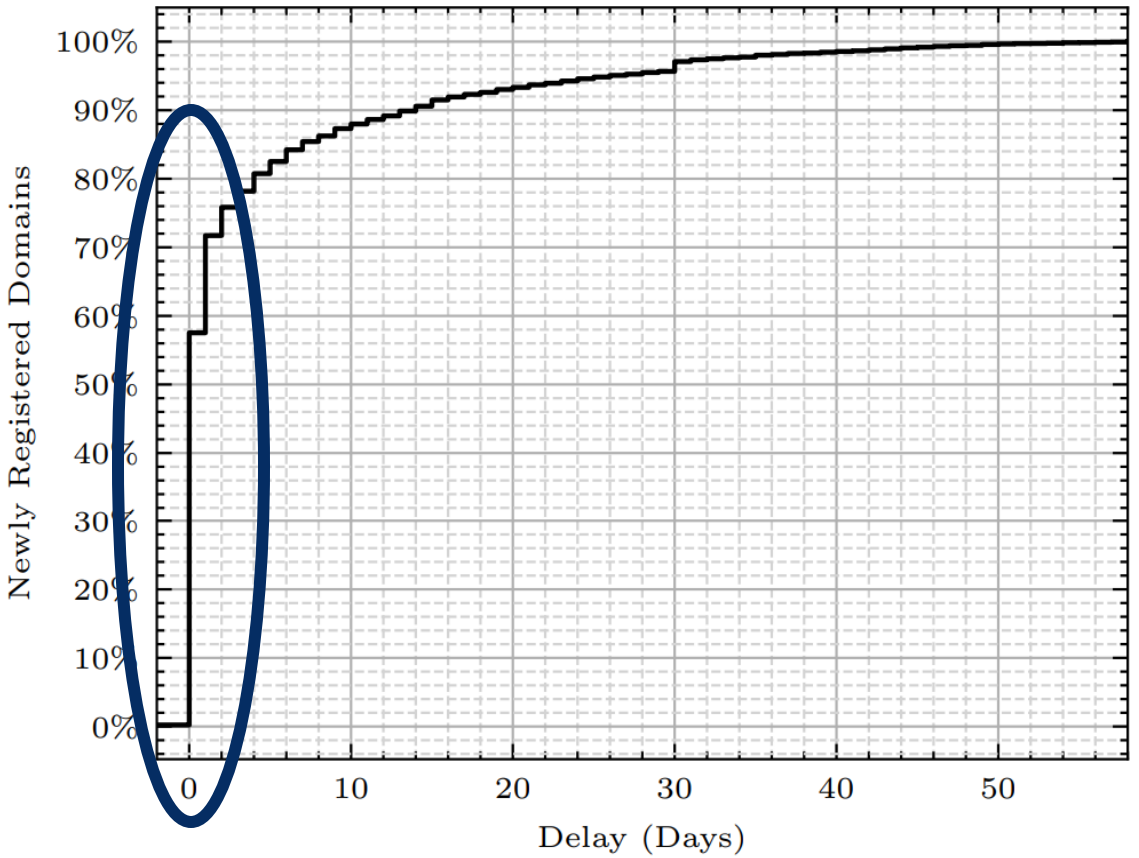
Poster: Through the ccTLD Looking Glass: Mining CT Logs for Fun, Profit and Domain Names

Authors: Raffaele Sommese and Mattijs Jonker | Authors Info & Claims

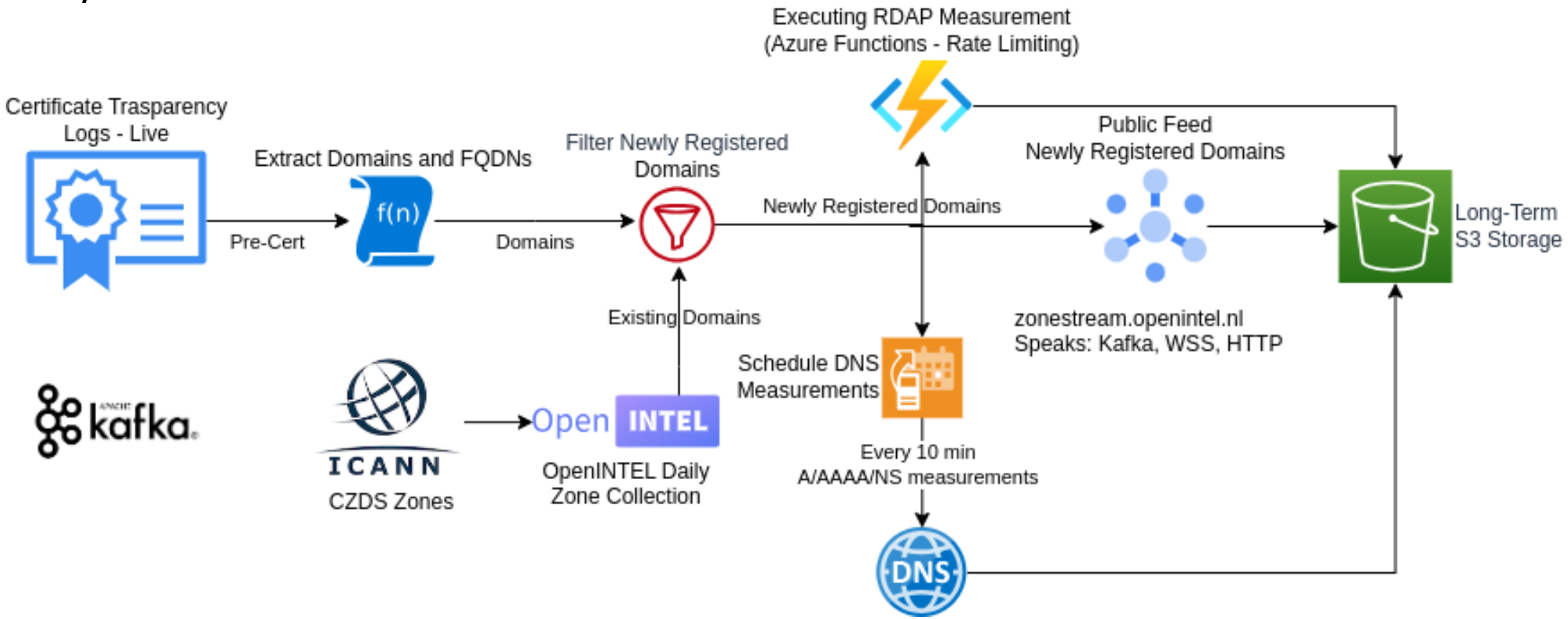
IMC '23: Proceedings of the 2023 ACM on Internet Measurement Conference • October 2023 • Pages 714 - 715
https://doi.org/10.1145/3618257.3624994

Published: 24 October 2023 Publication History

Check for updates



System Architecture



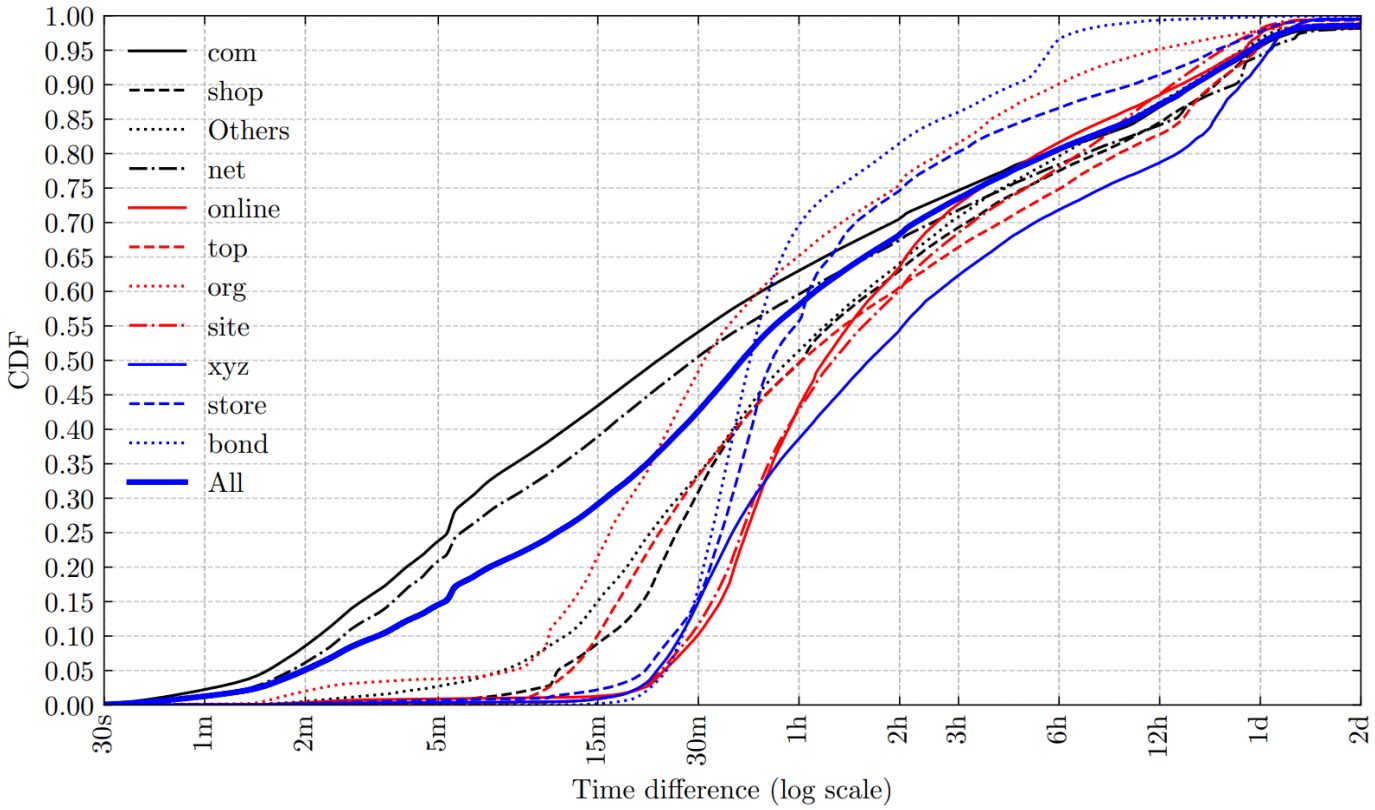
Newly registered domains

- We detect 42% of newly registered domains **before** they show up in the CZDS snapshot.
- ~76K domains per day.
- Almost 1 domain per second.
- 1% of newly registered domains **never** shows up in the next CZDS snapshot!

Detection Timeliness

Compared to RDAP registration date:

- 50% of domains detected **within 45 minutes** of their existence.
- ≈ 30% within 15 min.
- Small percentage of misclassified domains (as newly registered)



In CT Logs, but never in CZDS

- Approximately 1% of the newly registered domains **never** shows up in CZDS.
- Two main (possible) reasons:
  - Certificates issued for expired domains.
  - Domain lasting less than two zone snapshot interval -> **Transient domains**
- We used RDAP data to distinguish this two cases, finding **42K transient** domains over a 3-month period.
- **Invisible** so far to researchers!

Transient Domains

- Transient domains last at maximum 24 hours, half of them died within their **first 6 hours of life**.
- There is very **few legitimate reasons** for this, most of this registrations are liked to malicious behavior as confirmed by prominent registrars.
- Reasons for early removal include **abuse**, account suspensions, or credit card fraud.
- **Blocklists** do not promptly or in some cases ever detect transient domains!

Similar to long-lived domains

- **Half** of these transient domains were using **Cloudflare** as DNS provider (i.e., for their authoritative nameservers) and ≈35% of them used Cloudflare as a CDN provider.

Registrar	Domains	%
GoDaddy	8213	19.39%
Hostinger	6418	15.2%
NameCheap	4195	9.9%
Squarespace	2820	6.7%
Public Domain Registry	2625	6.2%
IONOS	2352	5.6%
Metaregistrar	1866	4.4%
NameSilo	1853	4.4%
Network Solutions, LLC	1670	3.9%
Tucows	1304	3.1%
Others	9042	21.3%
Total	42358	-

Table 3: Transient Domains Registrars Distribution.

How many transient domains we miss?

- We compared our transient domain feed with:
  - EPP transaction logs from .nl (ground truth, same idea of Rapid Zone Updates of .com, **non-public**)
  - Newly Observed Domains from DomainTools (passive DNS, **non-public**)

Our detection vs EPP Ground Truth

- In 3 months, .nl registry observed 714 domain names that were deleted in less than 24 hours in their registration system.
- Of those domains, 334 were registered and deleted such that they were never captured in zone file snapshots.
- With our methodology, we found only 99 transient domains, or **29.6%** of the 334 .nl-identified transient domain names over a period of 3 months.
- Researchers still have a huge blind spot regarding intra-day events.

Our detection vs DomainTools NOD

- DomainTools NOD feed detects in absolute numbers roughly 5% more domains than our methododology.
- However, the overlap between the two data sources is only ≈60%.
- This means that both methodologies provide an **additional**, although **not complete**, visibility into transient domains!

Where next?

- Transient domains indicates measure of **success** in registrars detecting malicious domains in their early stages, before they can do damage.
- However, each registrar has to independently **relearn** the same signals as threat actors move across different registrars to evade detection.
- In the meantime, transient domains, in which malicious activity dominate, have been **invisible to researchers**.

Time to resurrect Rapid Zone Updates

- “promote security and stability by providing a useful tool to online security companies, ISPs, search engines, financial services companies, and other stakeholders.”
- Due to the **ineffectiveness** of existing **uncoordinated** countermeasures, and the limited obligations of registrars to mitigate harm, we see a **need to expand transparency.**
- We can learn from history how to **mitigate the risk of abuse** of sharing data.
- CZDS represent a testament to the ability of managing this risks.

zonestream.openintel.nl