Associated Domain Discovery

SSAC Lightning Talk ICANN 84

26th October 2025

Motivation

- Attackers often **register domains in bulk** to support large-scale malicious campaigns (Vissers et al., 2017, Desmet, 2021) including phishing, malware delivery & botnet C2.
- Prior work hypothesizes a **preference for APIs or bulk registration tools** to streamline and automate these registrations (Nosyk, 2025).
- Use of APIs introduces **detectable patterns** *e.g.*, bursty, time-bound batches of registrations.
- Patterns may reveal attacker infrastructure **early in the attack lifecycle**.
- Understanding these patterns enables more **effective detection, intervention, and mitigation**.

Some Eyeballed Domain Patterns

We often see what look like related domains, *e.g.* (all in the same TLD):

```
estaxfelq.TLD
estaxfelr.TLD
estaxfelt.TLD
…
```

```
parcels1.TLD
parcels2.TLD
parcels3.TLD
…
```

Another recent example, less immediately obvious:

```
bitorderc.TLD
bitordern.TLD
bitswapf.TLD
bitswapn.TLD
bittaskq.TLD
bittaskz.TLD
blockorderw.TLD
```

```
blocktaskc.TLD
blocktasks.TLD
coinorders.TLD
coinorderw.TLD
coinswapf.TLD
coinswapk.TLD
…
```

Research Work

- Take seed domains from our RBLs
  - Spamhaus, SURBL, Phishtank, Urlscan, APWG, Urlhaus
- Use gTLD Registration Data
  - Bulk registration data access (BRDA data)
  - We do not have registrant or account data
- Newly Observed Domain Data
  - From URLScan.io

Currently evaluating three methods…

Discovery Method 1

Shared registration features (time-bound)

- Domains registered *close in time*
  - density based clustering (DBSCAN)
- Using the same registrar-nameserver combination
  - remember, we don't have registrant data
- Sharing similar lexical / string characteristics
  - Post clustering filter

Datasources: RBL feeds and BRDA data

Discovery Method 2

Shared registration features (loosely time-bound)

- Domains registered *over a longer time window*
  - Still using a temporal element
- Through the same registrar-nameserver combination
  - we still don't have registrant data
- Sharing similar lexical / string characteristics
  - random forest

Datasources: RBL feeds and BRDA data

Discovery Method 3

Shared infrastructure

- Domains sharing common IP-address (v4)
- and Nameserver combinations
  - but can be different registrar
- Sharing similar lexical / string characteristics
  - random forest

Datasources: RBL feeds and Newly Observed DNS resolution data

Note: CDNs, *etc.* are filtered

Measuring Similarity

Many features, both temporal and lexical, fed into a random forest classifier trained on manually tagged associated domains.

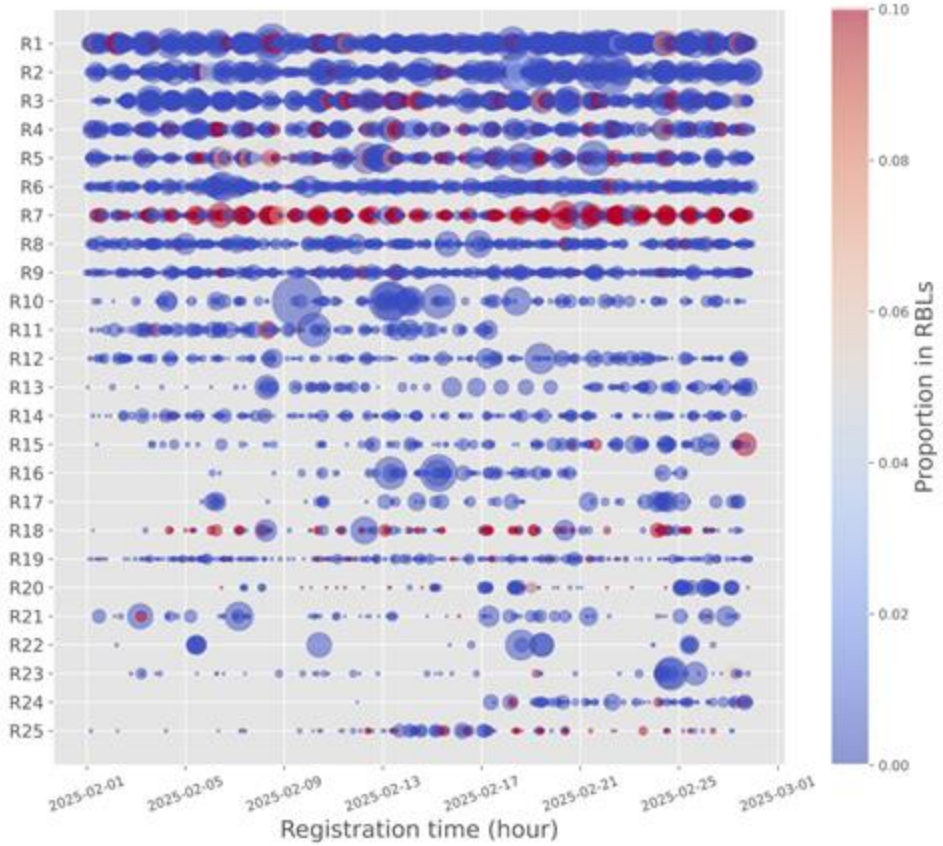
Lexical features include:

- String length
- Number of words
- Number of distinct characters
- Digit:letter ratio
- Levenshtein score
- Hyphen ratio

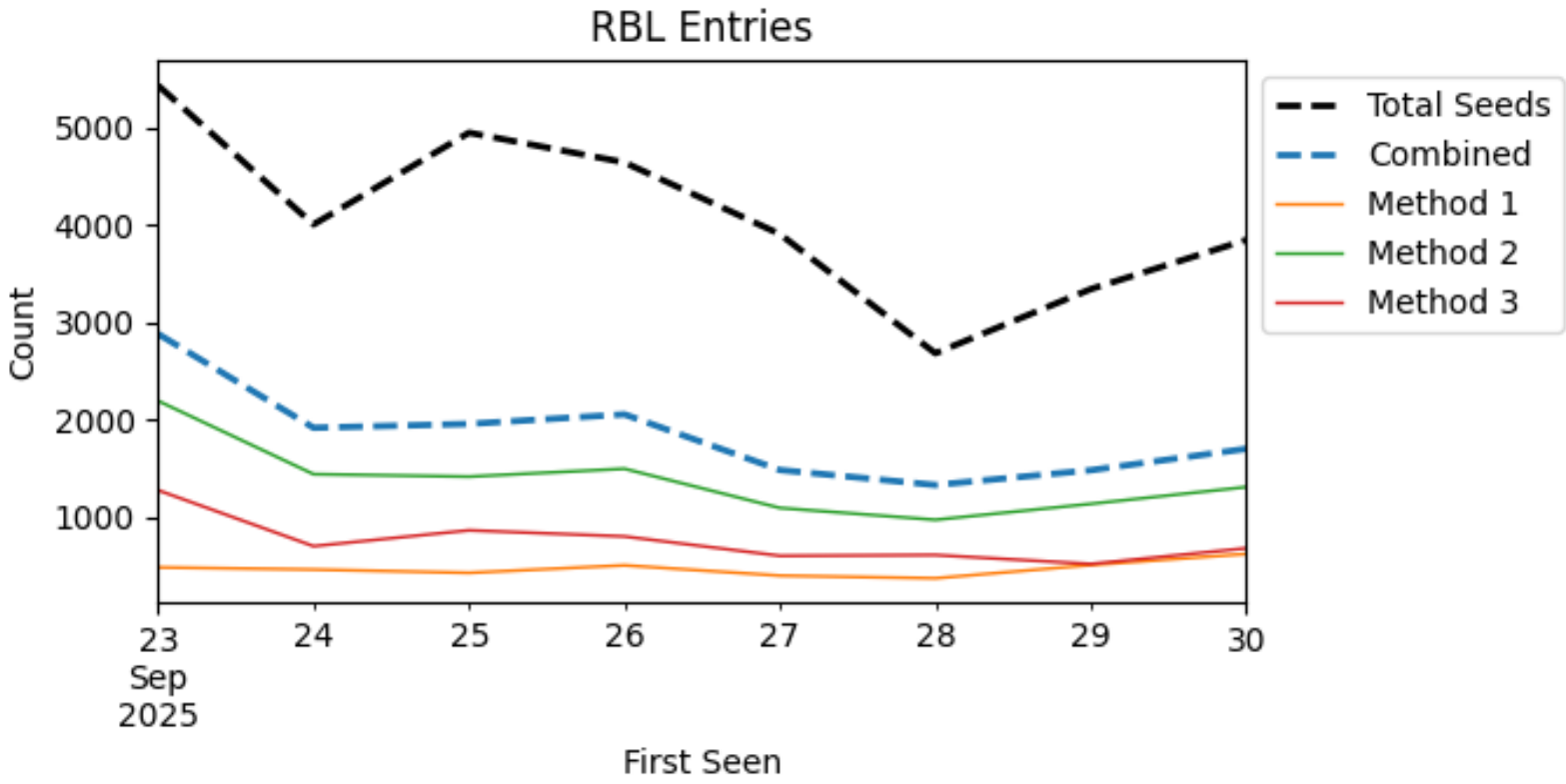
Temporal features include:

- First seen delta
- Registration density
  at different timescales
  (1 minute to 24 hours)

High level analysis – “batch” analysis



High level analysis – RBL Entries



Q&A

Full presentation by Sam Cheadle at:
APWG eCrime 2025, November 4th - 7th San Diego

T. Vissers *et al.*, 'Exploring the Ecosystem of Malicious Domain Registrations in the .eu TLD', *Research in Attacks, Intrusions, and Defenses*, vol. 10453, M. Dacier, M. Bailey, M. Polychronakis, and M. Antonakakis, Eds., in Lecture Notes in Computer Science, vol. 10453. , Cham: Springer International Publishing, 2017, pp. 472–493. doi: 10.1007/978-3-319-66332-6_21

Lieven Desmet, Jan Spooren, Thomas Vissers, Peter Janssen, and Wouter Joosen. 2021. Premadoma: An Operational Solution to Prevent Malicious Domain Name Registrations in the .eu TLD. Digital Threats 2, 1, Article 2 (March 2021), 24 pages. https://doi.org/10.1145/3419476

Y. Nosyk, M. Korczynski, S. Maroofi, J. Bayer, Z. Odgerel, A. Duda, S. Tajalizadehkhoob, and C. Gañán, 'INFERMAL: Inferential analysis of maliciously registered domains'

ICANN Webpage and Social Media Links


icann.org


@icann


facebook.com/icannorg


youtube.com/icannnews


flickr.com/icann


linkedin.com/company/icann


instagram.com/icannorg