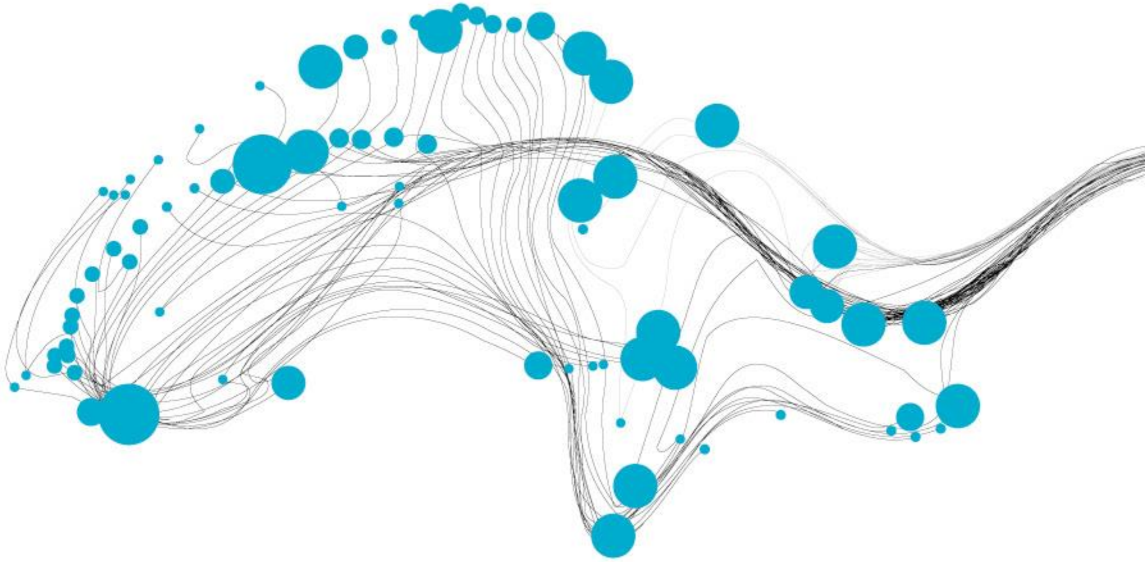
Unveiling Domain Blocklist Performance:

An Analysis over Four Years



Antonia Affinito*, Alessio Botta+, Anna Sperotto*

*University of Twente, the Netherlands

+University of Napoli “Federico II”, Italy

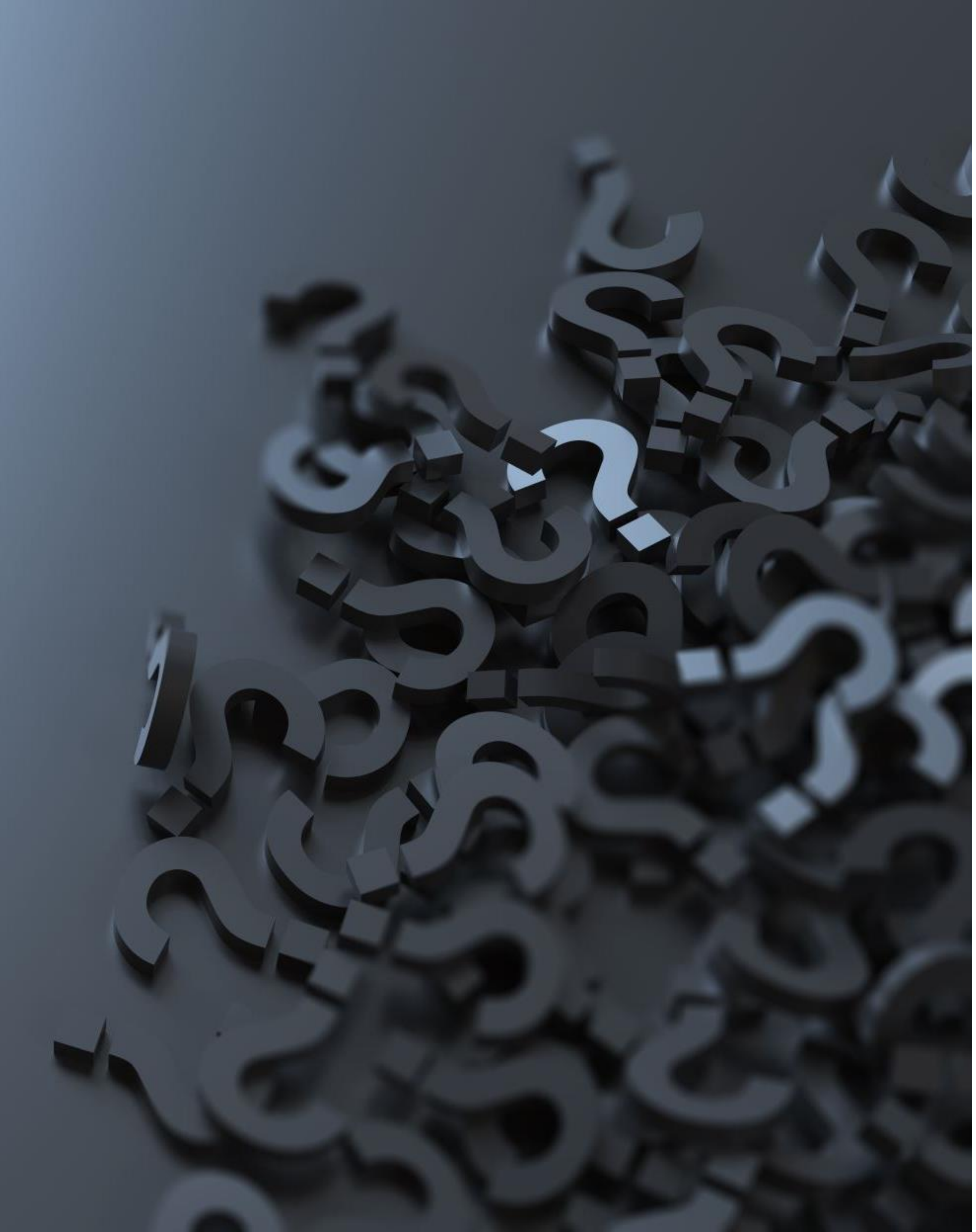
UNIVERSITY
OF TWENTE.

Domain Blocklists

- Domain blocklists are lists of **malicious** or **suspicious** domains used to protect against cyber threats
- Maintained by a variety of entities
  - Private **organizations**, **research institutions**, and **individual contributors**

Problem Statement

- Blocklists differ in **detection speed**, **update frequency**, and **domain overlap**
- Despite widespread use, blocklists' effectiveness and long-term consistency remain to be **explored**



Objective

- Provide a characterization of domain blocklists' **effectiveness** and **evolution over a four year-period**
  - Evaluate the update **frequency** of domain entries in blocklists over time
    - Using external validation source (i.e., DNS.coffee)
  - Identify prominent **characteristics** in the blocklists

Analyzed Blocklists

- We analyzed 13 **domain blocklists**
  - 2020-03-16 to 2024-08-07, covering a total of more than **4 years**

Blocklist Name
Spamhaus DBL
PhishTank
Cybercrimetracker
Tolouse {DDoS, Malware, Crypto, Phishing}
Digitalside
OpenPhish
Phishingarmy
Vxvault
Ponmocup
Quidsup

Overview of Related Work

- Study of the transparency and dynamics of **various open** blocklists [1, 2]
- **Honeypots** are used to observe interactions and assess whether domains are being targeted for malicious activity [1]
- Analysis of the overlap among blocklists within the **same** category to understand redundancy and coverage [3,4]
- Assessment of how frequently blocklists are updated [1,2,3]

[1]: Á. Feal et al., "Blocklist Babel: On the Transparency and Dynamics of Open Source Blocklisting" in IEEE Transactions on Network and Service Management, 2021
[2]: M. Umizaki et al. "Understanding the Characteristics of Public Blocklist Providers," IEEE Symposium on Computers and Communications (ISCC), 2022
[3]: S. Bell et al. “An Analysis of Phishing Blacklists: Google Safe Browsing, OpenPhish, and PhishTank”. In Proceedings of the Australasian Computer Science Week Multiconference, 2020
[4]: Velden, J. van der “Blacklist, do you copy? Characterizing information flow in public domain blacklists”

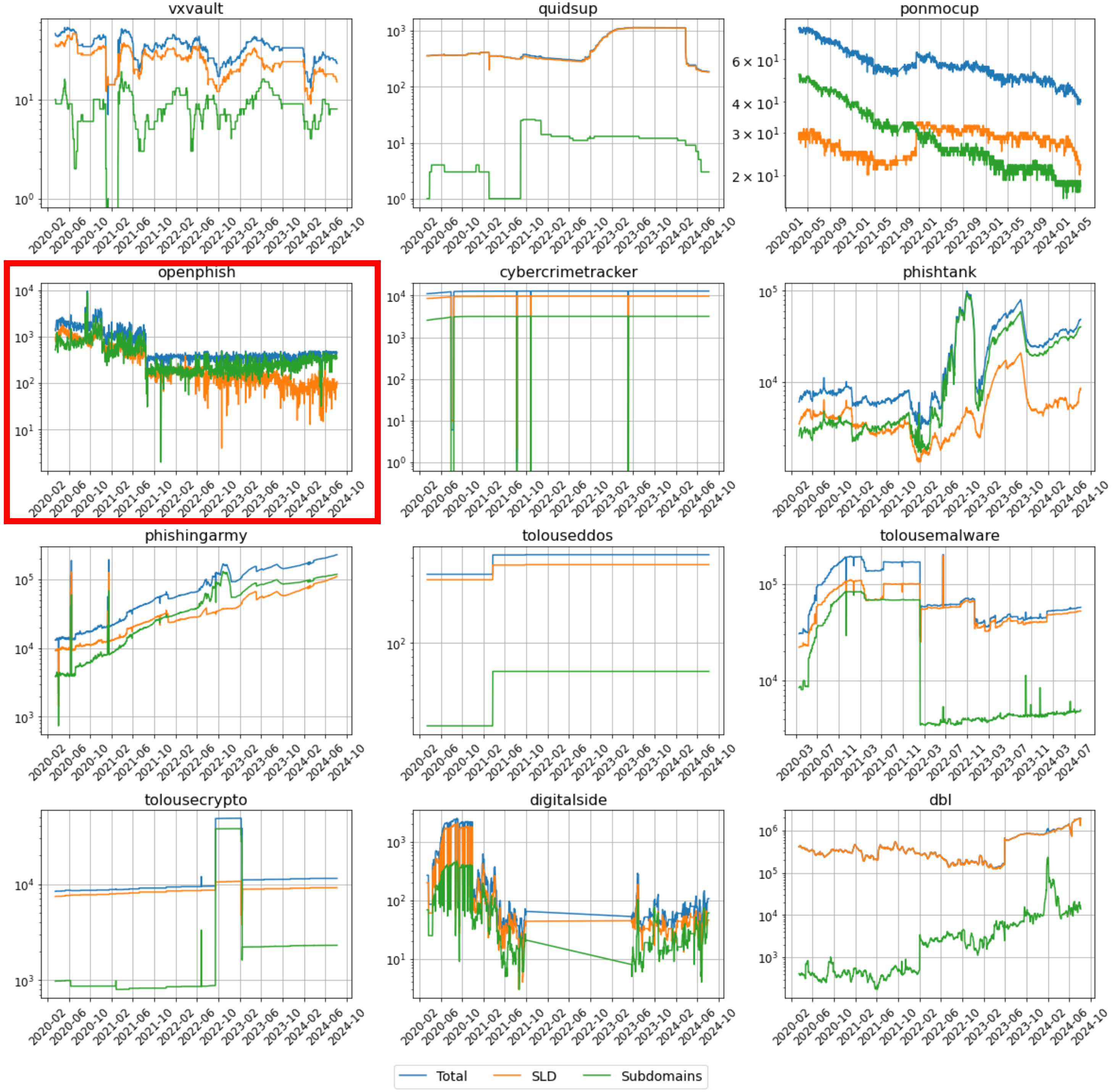
What Sets Our Study Apart

- **Extendend Analysis**: Examined domain blocklists from 2020 to 2024 to capture long-term trends
- **Update Frequency**: Used DNS data (dns.coffee API) to identify outdated entries and assess domain validity

PRELIMINARY RESULTS

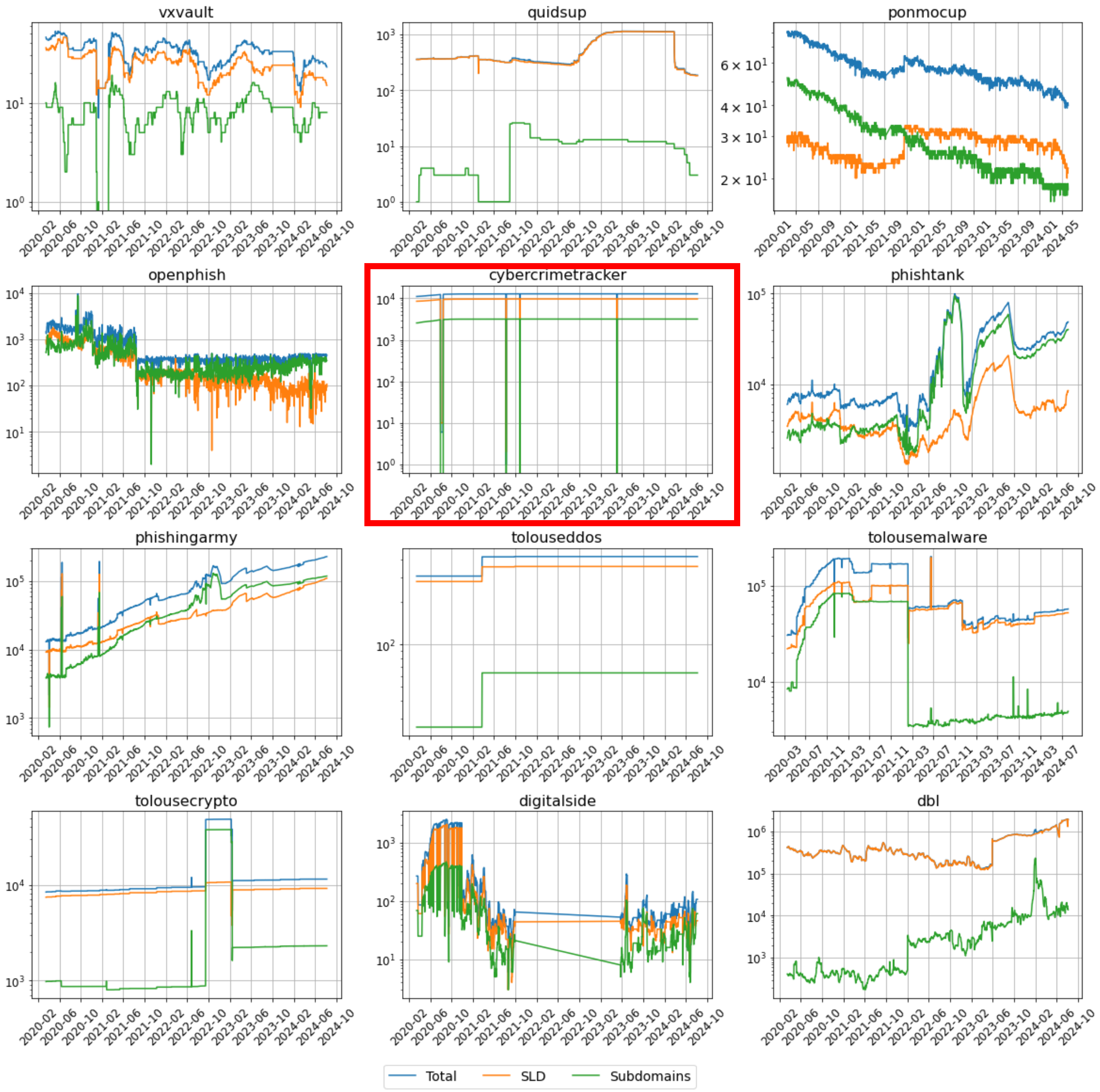
Number of Domains Over Time

- **Decrease** in OpenPhish starting mid-2021
- **Consistent** number of domains in Cybercrimetracker over time
- **Rising** domain counts in PhishTank, Phishingarmy, and DBL over time
- **More** SLD than FQDN in all blocklists, except Ponmocup and OpenPhish



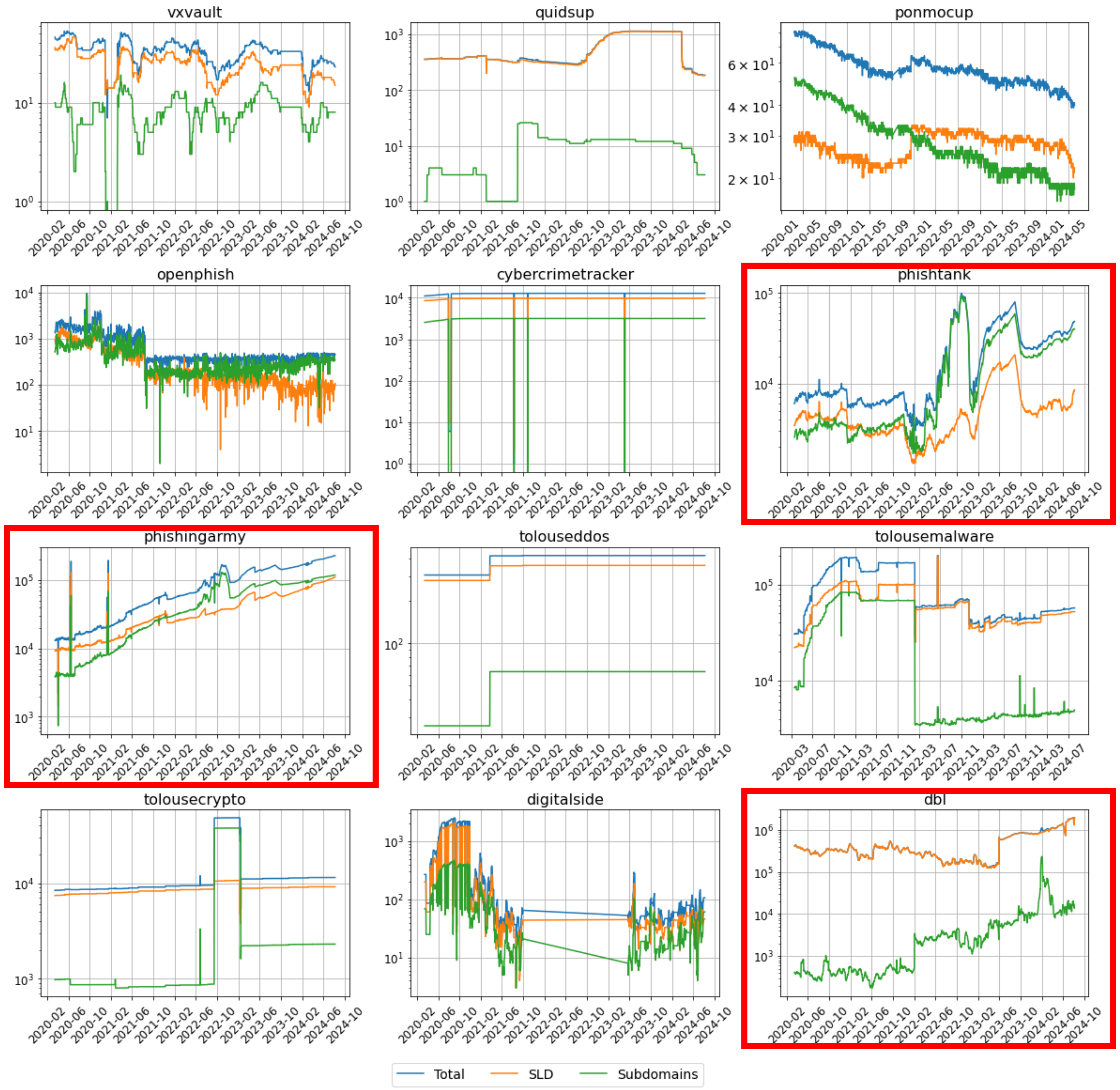
Number of Domains Over Time

- **Decrease** in OpenPhish starting mid-2021
- **Consistent** number of domains in Cybercrimetracker over time
- **Rising** domain counts in PhishTank, Phishingarmy, and DBL over time
- **More** SLD than FQDN in all blocklists, except Ponmocup and OpenPhish



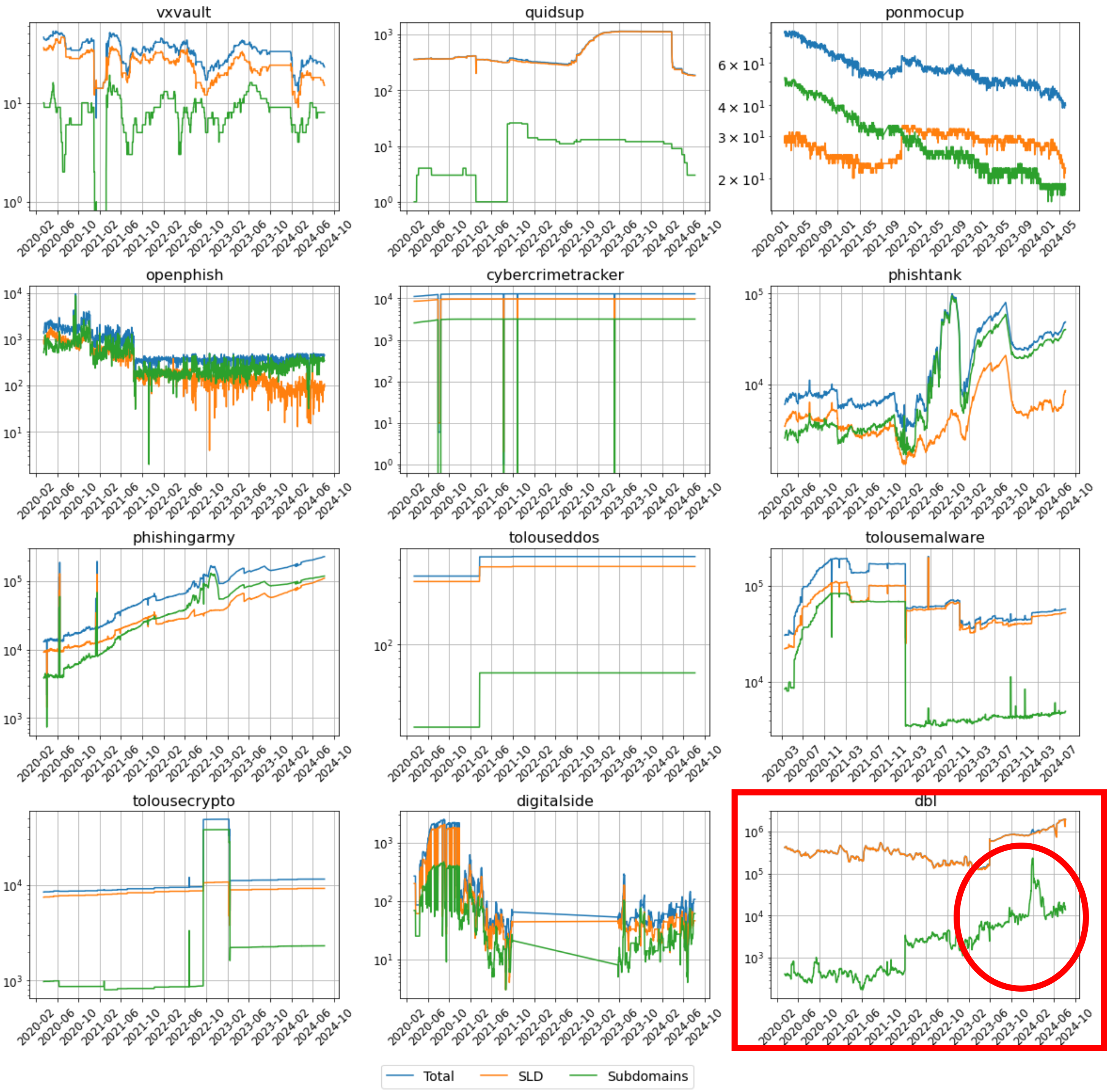
Number of Domains Over Time

- **Decrease** in OpenPhish starting mid-2021
- **Consistent** number of domains in Cybercrimetracker over time
- **Rising** domain counts in PhishTank, Phishingarmy, and DBL over time
- **More** SLD than FQDN in all blocklists, except Ponmocup and OpenPhish



Number of Domains Over Time

- **Decrease** in OpenPhish starting mid-2021
- **Consistent** number of domains in Cybercrimetracker over time
- **Rising** domain counts in PhishTank, Phishingarmy, and DBL over time
- **More** SLD than FQDN in all blocklists, except Ponmocup and OpenPhish

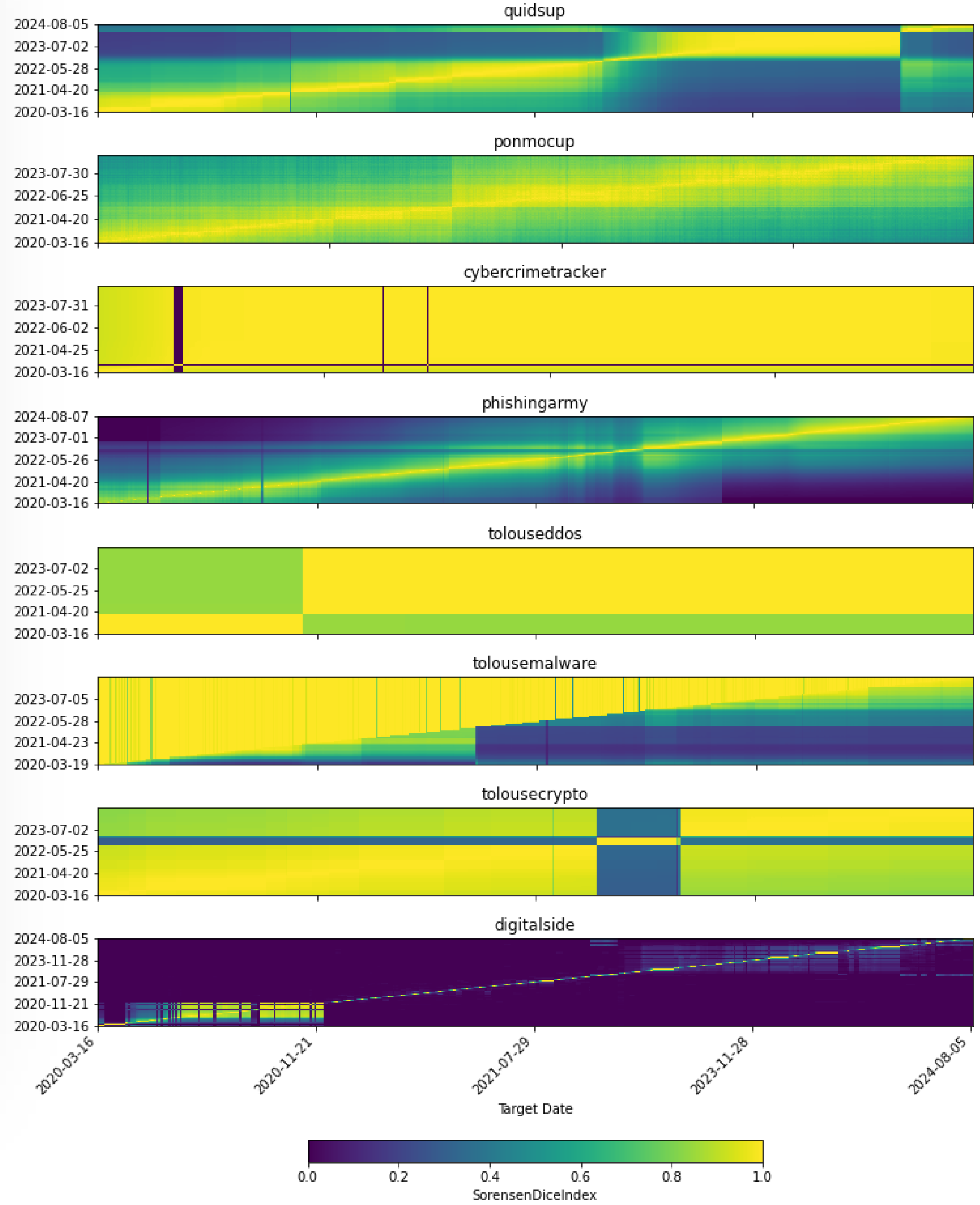


Update Frequency Over Time (1/2)

- **Variable Update Frequencies**: Some blocklists (e.g., digitalside, phishingarmy) exhibit small intervals with a low data variation

- **Constant Data Values**: Some blocklists (e.g., cybercrimetracker, tolouse) show a high Sørensen-Dice coefficient, indicating they rarely change over time

Sørensen-Dice Coefficient: similarity of collected domains over time

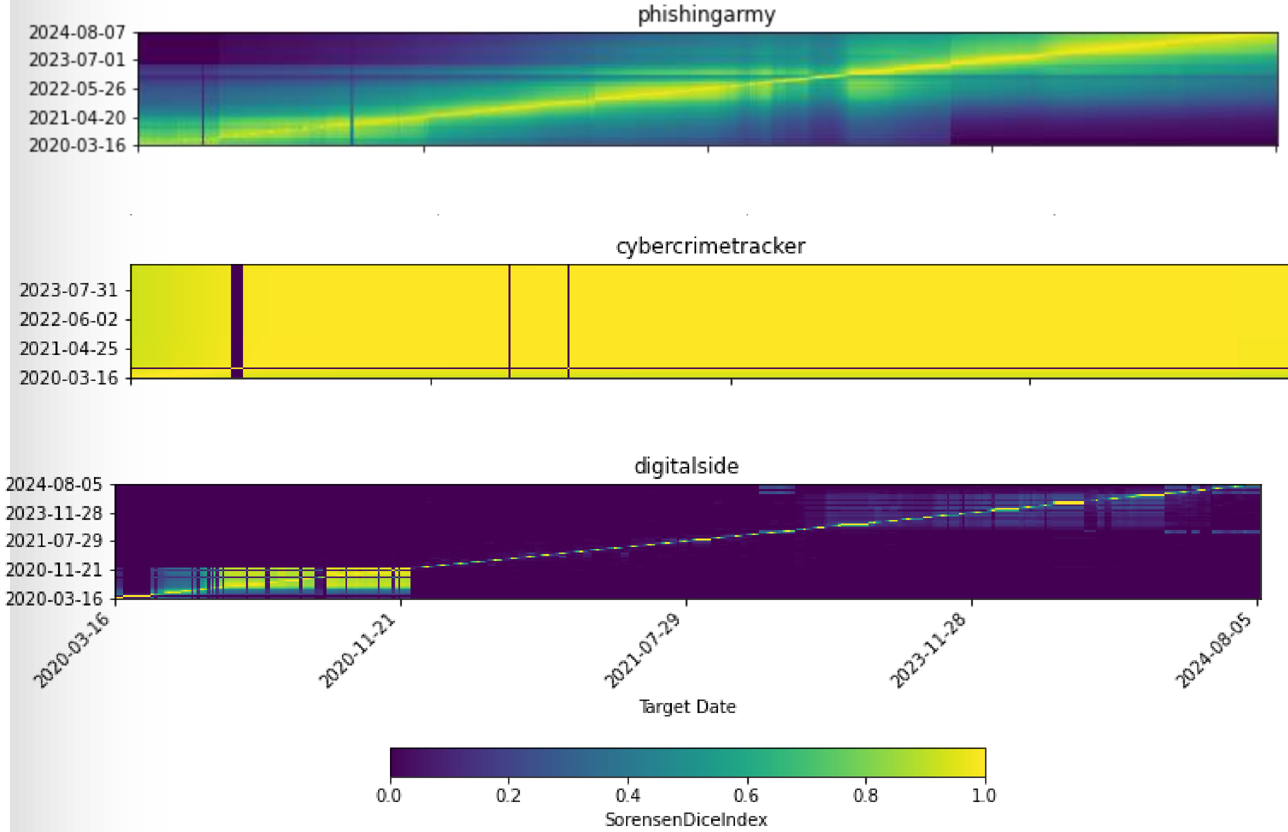


Update Frequency Over Time (1/2)

- **Variable Update Frequencies**: Some blocklists (e.g., digitalside, phishingarmy) exhibit small intervals with a low data variation

- **Constant Data Values**: Some blocklists (e.g., cybercrimetracker, tolouse) show a high Sørensen-Dice coefficient, indicating they rarely change over time

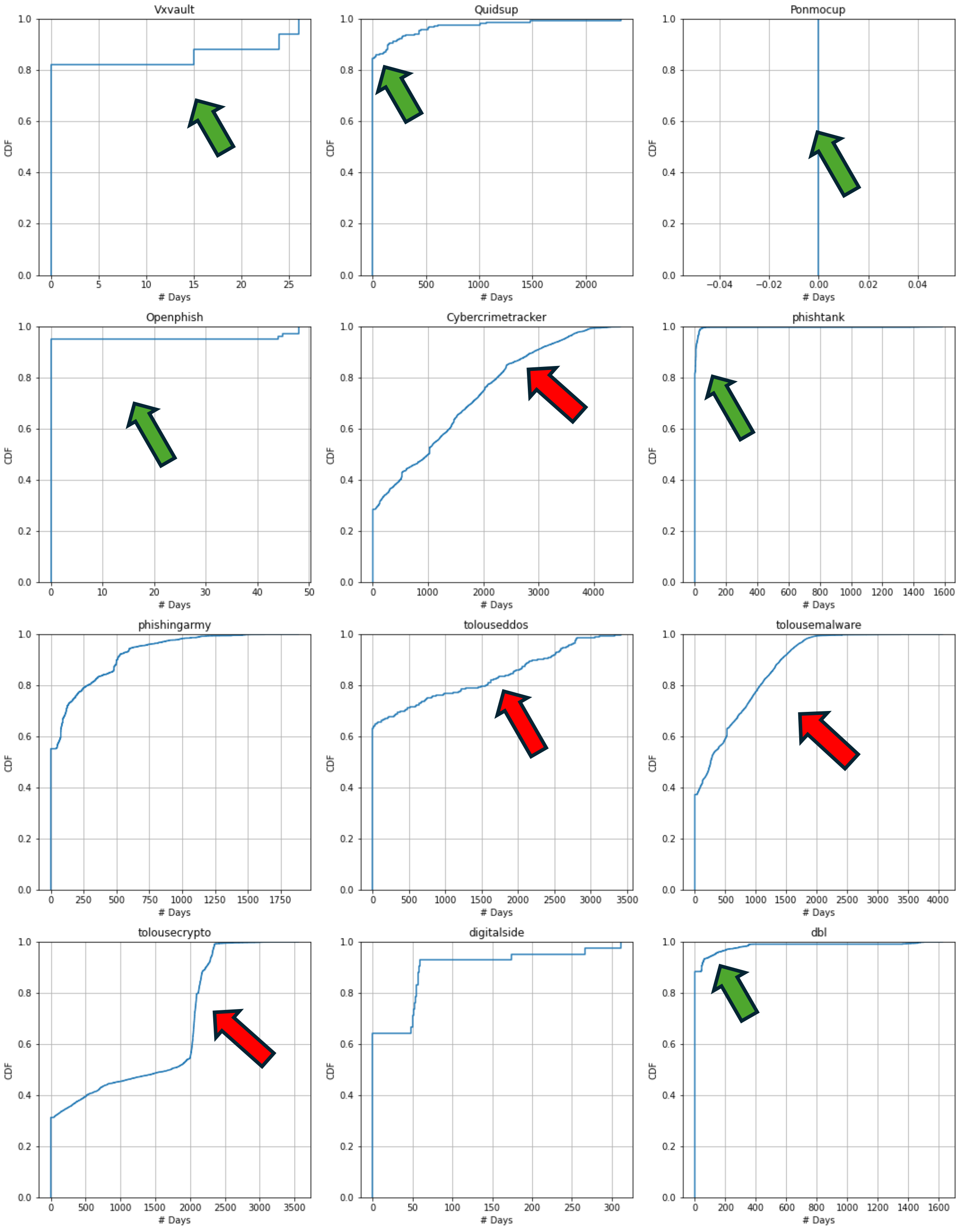
Sørensen-Dice Coefficient: similarity of collected domains over time



Update Frequency Over Time (2/2)

- Red: **higher probablity** of including domains that have a expiration date **back than 500 days reaching 4000 days**

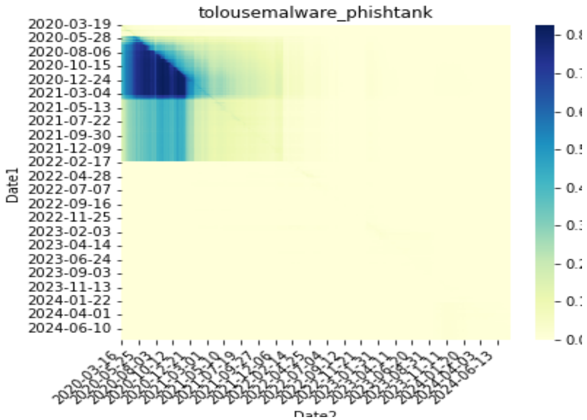
- Green: **higher probablity** of including **active domains**

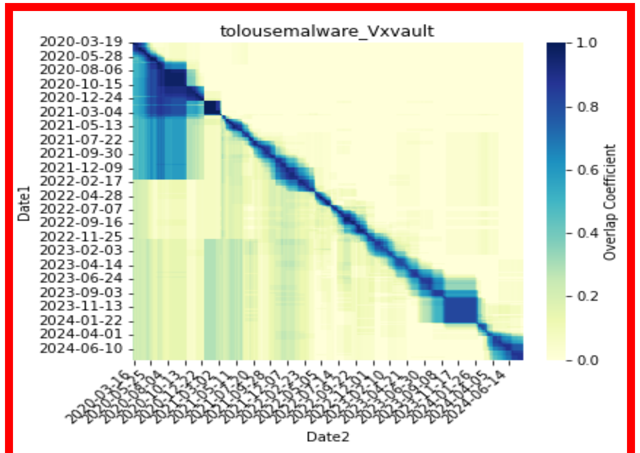


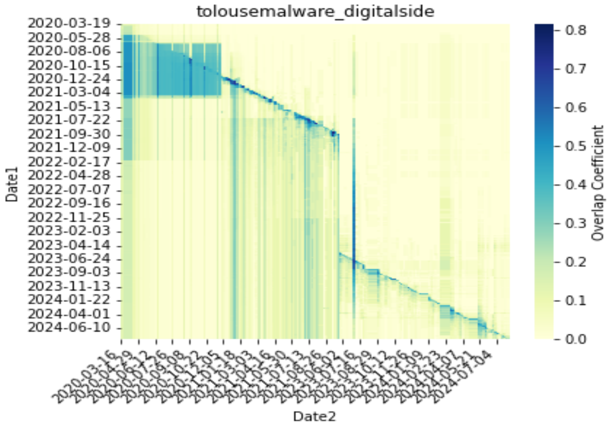
Number of days since the expiration dates of the domain names

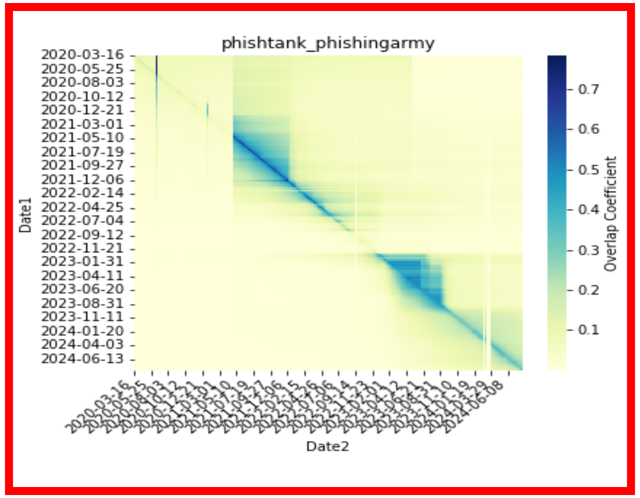
Blocklist Overlap

- **Overlap Coefficient**
- Overlap coefficient is very **low**
    - Only **16.7%** of combinations exceed 0.5





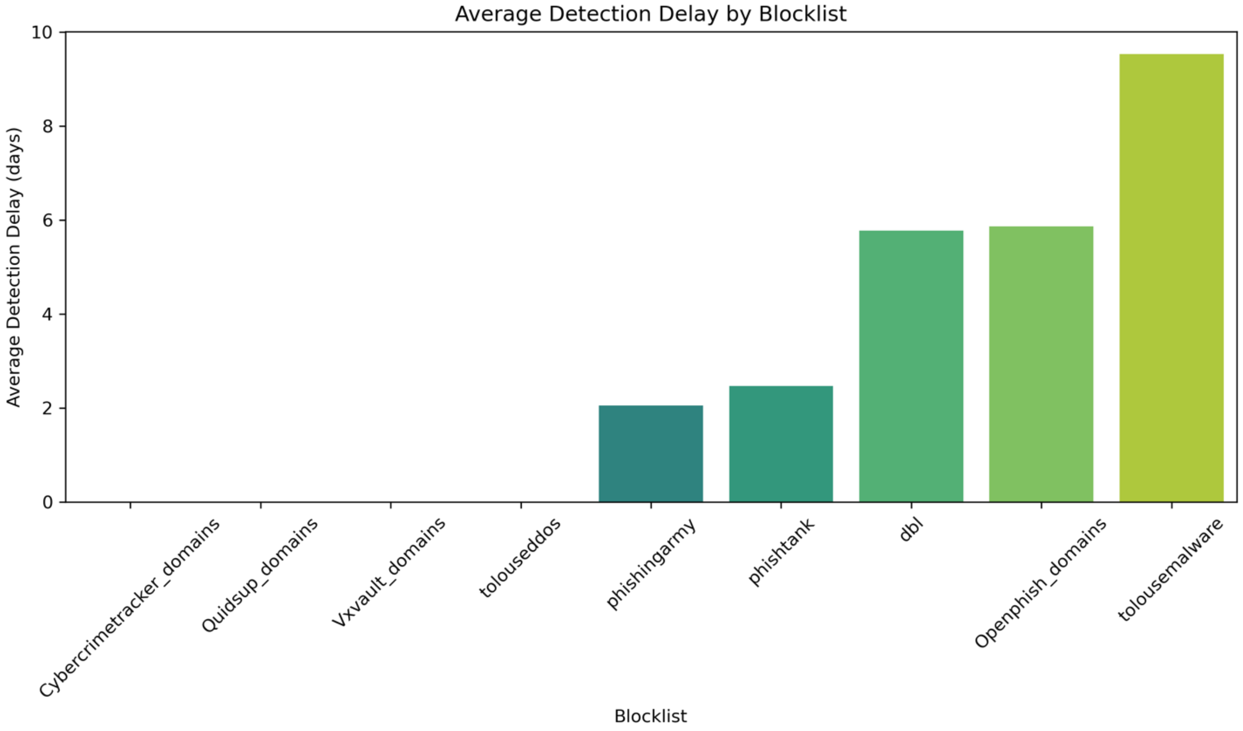




Detection Delay by Blocklist

Examples:

- zzjtjgur1.duckdns.org,**phishtank,2024-07-02**,0.0; **dbl,2024-07-12**,10.0
- zlmbrasharefile.com,**dbl,2024-07-12**,0.0;**phishtank,2024-07-23**,11.0
- ygdkb-f2f2b7.ingress-daribow.ewp.live,**Openphish,2024-07-11**,0.0;**phishingarmy,2024-07-18**,7.0



Conclusion

- Blocklists **vary** in **update** frequency, with some tending to list old expired domain names (up to 4000 days past expiration)

- The **low overlap** suggests that each blocklist captures distinct sets of threats, emphasizing the importance of using multiple lists for comprehensive security.

Thanks for your attention

If you have any questions

contact: a.affinito@utwente.nl