

Building a Resilient Domain Whitelist to Enhance Phishing Blocklist Accuracy

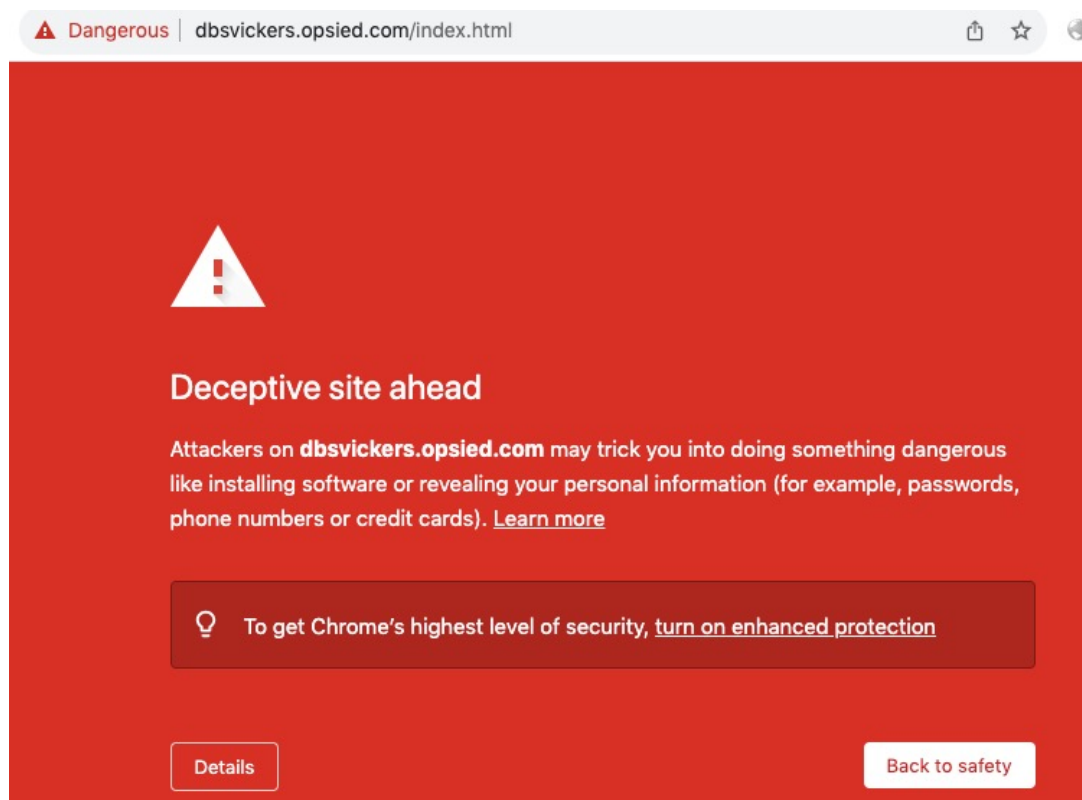
Maciej Korczyński, Sourena Maroofi, Jan Bayer, Andrzej Duda

Grenoble Institute of Technology, Université Grenoble Alpes
maciej.korczynski@univ-grenoble-alpes.fr

23 October 2023
ICANN Tech Day

Importance of blocklists

- URL and domain blocklists are crucial in phishing prevention, malware protection, reputation management, etc...



Motivation: problems with blocklists

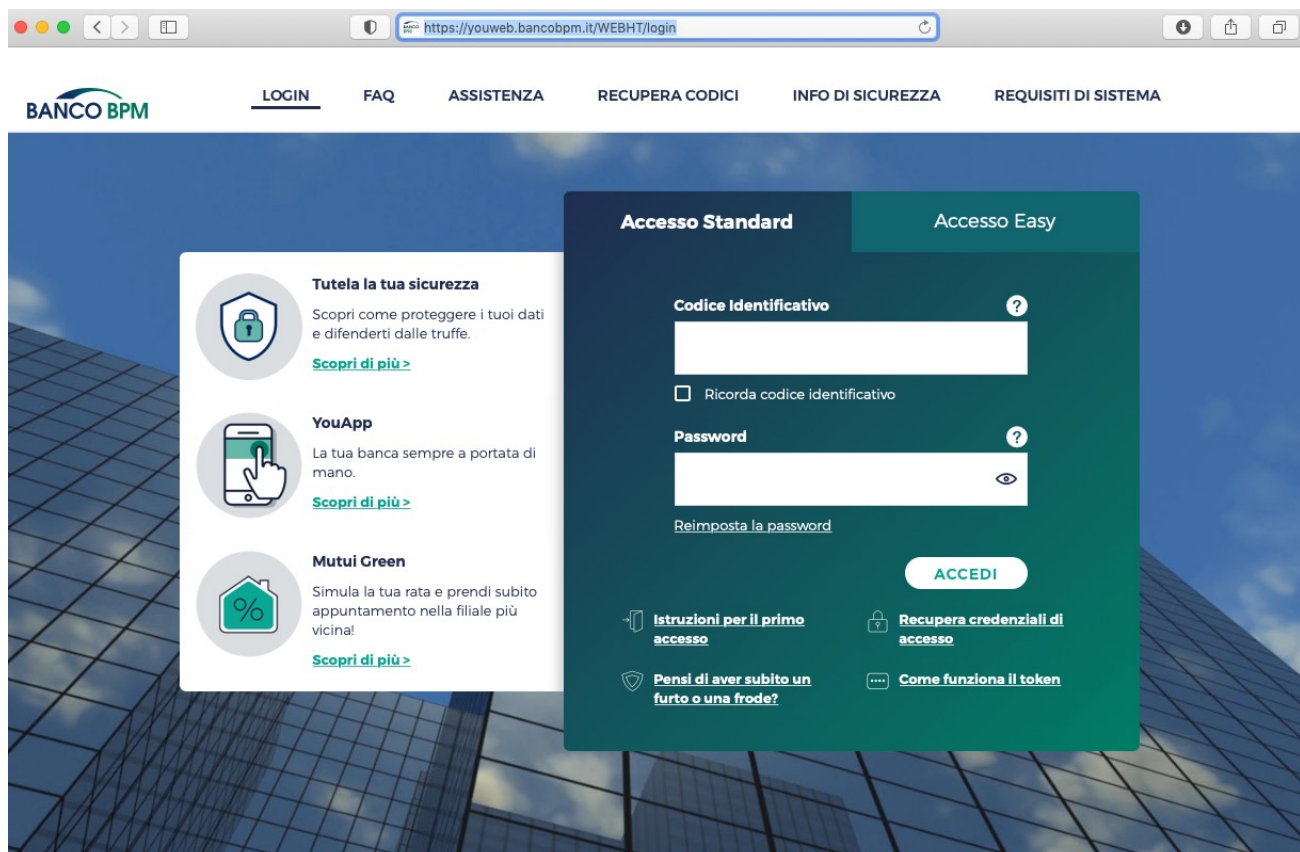
- Blocklists (may) contain
 1. False positives (i.e., URLs and domain names incorrectly categorized as malicious)
 2. Phishing simulation URLs and domain names
 3. Malicious URLs associated with special services' domain names

Agenda

- Problem statement
- Consequences of false positives, phishing simulations, and special domain names on blocklists
- Proposed solutions
- Resilient domain whitelist to enhance phishing blocklist accuracy

False positives: example

- URLs and domain names incorrectly categorized as malicious
Blocklisted URL: <https://youweb.bancobpm.it/WEBHT/login>



False positives: example

- URLs and domain names incorrectly categorized as malicious

```
whois bancobpm.it
```

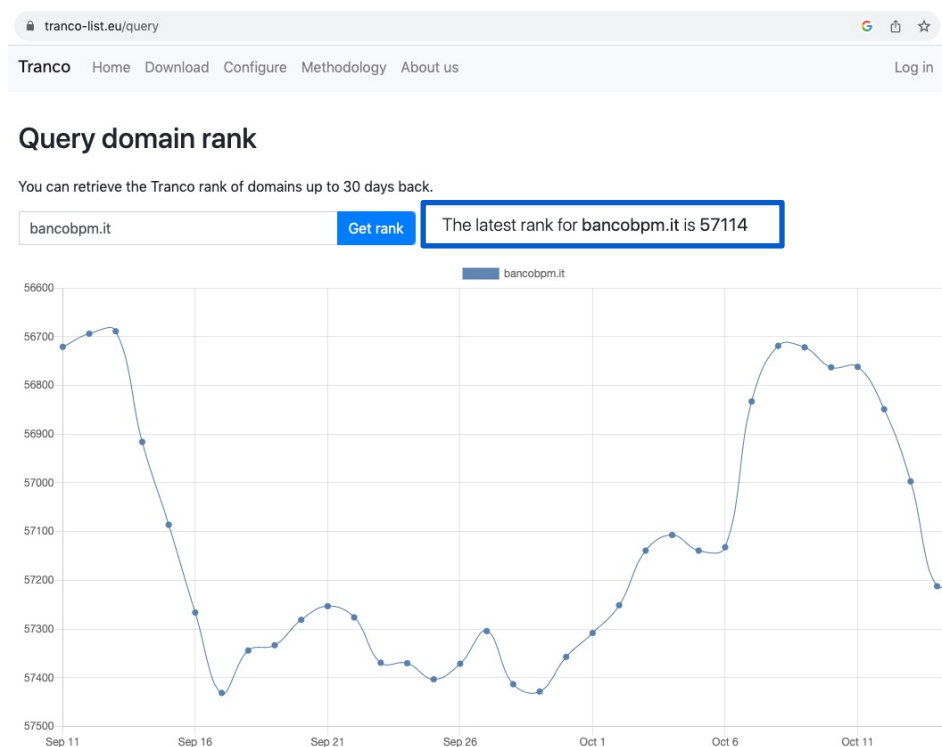
```
Created: 2016-03-28
```

False positives: example

- URLs and domain names incorrectly categorized as malicious

whois bancobpm.it

Created: 2016-03-28



False positives: example

- URLs and domain names incorrectly categorized as malicious

whois bancobpm.it

Created: 2016-03-28

Certificate Viewer: youweb.bancobpm.it

General Details

Issued To

Common Name (CN) youweb.bancobpm.it
Organisation (O) BANCO BPM SOCIETA' PER AZIONI
Organisational Unit (OU) <Not part of certificate>

Issued By

Common Name (CN) Sectigo RSA Extended Validation Secure Server CA
Organisation (O) Sectigo Limited
Organisational Unit (OU) <Not part of certificate>

Extended Validation (EV) certificates requires verification of the requesting entity's legal identity prior to issuance



Consequences of false positives

- Users may not be able to access benign websites and services
- Diminished trust in services provided under blocklisted domains
- Decreased trust in intermediaries responsible for domain registration, hosting, and blocklists
- Notifications to intermediaries take time and need to be verified
- Legitimate domain names taken down

Phishing simulation: example

- Phishing simulation involves registering a domain name and setting up a website that mimics the organization's brand

Blocklisted URL:

<http://login.safebank.onl/s/452/ac3762/cfba4963-f930-4006-b4cf-6cb4af083560>

PDF Online English

File Verification for Adobe PDF

☒ View File Online
☐ Download File
☐ Send File to Email

Username

Password

Login To View File

To access our online secure documents page, you are required to login your email address. This is to ensure you are the rightful recipient for the protected file. Unauthorized access is prohibited.

Phishing simulation: example

- Phishing simulation involves registering a domain name and setting up a website that mimics the organization's brand

Blocklisting date: 2023-07-29

`whois safebank.onl`

Creation Date: 2015-07-17

Phishing simulation: example

- Phishing simulation involves registering a domain name and setting up a website that mimics the organization's brand

Blocklisting date: 2023-07-29

whois safebank.onl

Creation Date: 2015-07-17

Registrar: MarkMonitor Inc.

Registrar IANA ID: 292

Phishing simulation: example

- Phishing simulation involves registering a domain name and setting up a website that mimics the organization's brand

Blocklisting date: 2023-07-29

whois safebank.onl

Creation Date: 2015-07-17

Registrar: MarkMonitor Inc.

Registrar IANA ID: 292

Web content of login.safebank.onl

How did I get here?

This was an authorized phishing simulation from your organization as part of your program.

Spear Phishing

Spear phishing messages target small groups or individuals. Attackers personalize these messages to bypass technical controls like spam filters.

Even if you receive an email that looks legitimate, always use caution. Keep these tips in mind to help spot a phish:

Keep your emotions in check.

Phishers frequently use emotions like fear, curiosity, and greed to compel recipients to open attachments or click links.



Look at the domain name.

Some attackers modify domains to catch targets off guard. For example, if the correct domain is www.example.com, the phishers may register "example.com" or "example.co".



Think twice.

Read emails thoroughly and be wary of words like Caution, Act Now, and Warning, which draw your attention and make you act quickly.



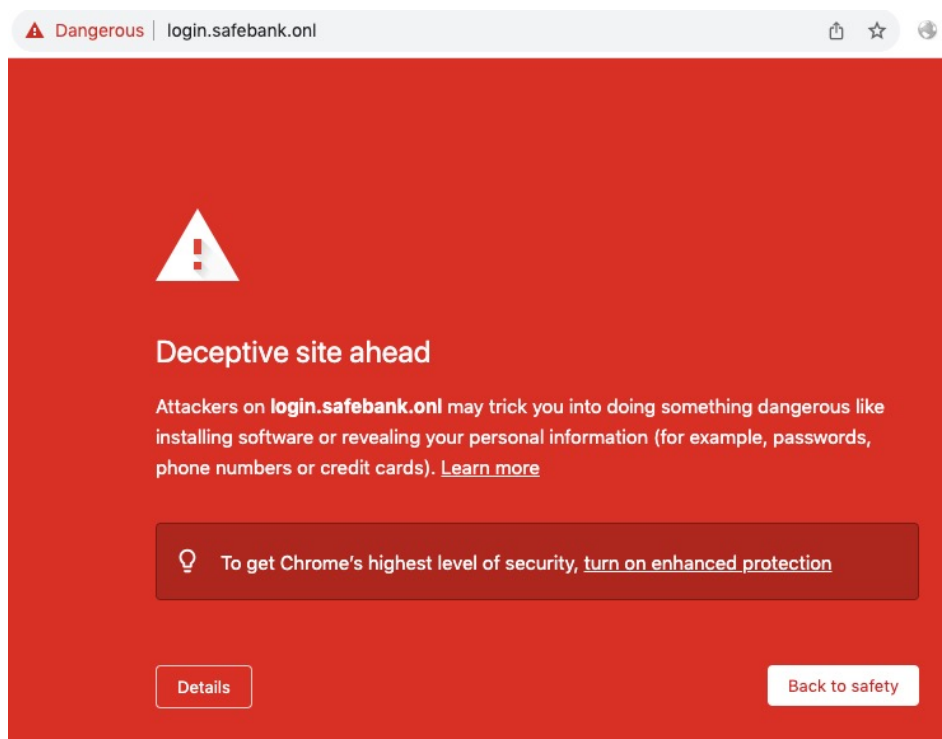
Always verify.

If you know the sender, follow up with a quick phone call. Report any suspicious emails.



Consequences of blocklisting phishing simulation websites and domain names

- Decreased trust in intermediaries responsible for domain registration and hosting
- Notifications to intermediaries take time and need to be verified



- Community service: Phishing Simulation Whitelist (PSW) allows verified parties to submit and whitelist domain names intended for use in phishing simulations or awareness campaigns within a specified validity period
- Usage: PSW accessible by everyone, will be used in DNSAI COMPASS



Phishing Simulation

Phishing simulations enable organizations to identify the most vulnerable areas within their organization to phishing threats. Typically, this involves registering a domain name and setting up a website that mimics the organization's brand. However, a challenge arises when such domain names or websites are detected by anti-phishing engines. This detection can have adverse consequences for the registrars and Top-Level Domain (TLD) registries associated with the registered domain names, impacting their reputation negatively.

To address this issue, the Phishing Simulation Whitelist (PSW) has been established as a community service. It allows verified parties to submit and whitelist domain names intended for use in phishing simulations or awareness campaigns within a specified validity period. The list of whitelisted domain names will be accessible to verified industry partners, including blocklist providers, and will be incorporated into DNSAI Compass reports. This ensures that these domains are excluded from analysis, safeguarding the reputation of intermediaries involved in domain registration and hosting.

How to Submit data

There are two possible ways to submit domain names to the whitelist: 1) using the API endpoint (described at the end of the page) or, 2) Using the provided online form. Both options need **access token** which is provided to verified partners by [contacting us](#).

The whitelist is published [here](#) every day (receives daily updates) in .CSV format.

Registered domain name	Date (yyyy-mm-dd)
Provided Token	
Submit	

API Documentation

You can use the following API endpoint to submit data automatically to our system. Here, we describe the parameters of the API.

Special domain names

- The domain registered by a benign organization
- Offers a range of services to users that could potentially be misused for the distribution of malicious content:
 - ✓ free subdomain providers (e.g., 000webhostapp.com, blogspot.ae) and paid subdomain providers (e.g., app.link),
 - ✓ dynamic DNS providers (ddns.net),
 - ✓ file sharing services (e.g., dropbox.com, github.com),
 - ✓ cloud service providers (amazonaws.com),
 - ✓ free page providers (e.g., linkedin.com),
 - ✓ text sharing services (e.g., pastebin.com),
 - ✓ link redirection services (rs6.net),
 - ✓ InterPlanetary File System (e.g., cloudflare-ipfs.com),
 - ✓ URL shorteners (e.g., bit.ly, tinyurl.com)

Special domains: example

↩ Reply

↩ Reply All ▾

➡ Forward

📁 Archive

🔥 Junk

🗑 Delete

More ▾

From DGS | Grenoble INP <noreply@grenoble-inp.fr> ☆

Subject **[SPAM HIGH]ATTENTION** 15:14

To maciej.korczynski@grenoble-inp.fr ★

Avertissement de boîte aux lettres!

Vous avez atteint la capacité de stockage de votre courrier et ne pouvez plus recevoir de nouveaux messages tant que vous n'avez pas augmenté la capacité de stockage.

Veuillez suivre le lien ci-dessous pour augmenter la capacité de stockage.

[ACCROÎTRE LA CAPACITÉ](#)

Cordialement Votre,
Directrice Générale des Services
Chef de l'administration
Institut polytechnique de Grenoble (Grenoble INP)
Institut d'ingénierie et de gestion de Grenoble
Université Grenoble Alpes
46 avenue Félix Viallet - 38031 Grenoble - France
www.grenoble-inp.fr
© Copyright 2020 | Grenoble Institute of Technology (Grenoble INP) - Université de Grenoble, France.

Grenoble



|



Institut polytechnique de Grenoble
(Grenoble INP Institute of Engineering and Management)
46 avenue Félix Viallet

99%

Unread: 2461

Total: 27698

 Today Pane ▾

Special domains: example

Reply

Reply All

Forward

Archive

Junk

Delete

More

From DGS | Grenoble INP <noreply@grenoble-inp.fr> ☆

Subject [SPAM HIGH ATTENTION] 15:14

To maciej.korczynski@grenoble-inp.fr ☆

Avertissement de boîte aux lettres!

Vous avez atteint la capacité de stockage de votre courrier et ne pouvez plus recevoir de nouveaux messages tant que vous n'avez pas augmenté la capacité de stockage.

Veuillez suivre le lien ci-dessous pour augmenter la capacité de stockage.

[ACCROÎTRE LA CAPACITÉ](#)

Cordialement Votre,
Directrice Générale des Services
Chef de l'administration
Institut polytechnique de Grenoble (Grenoble INP)
Institut d'ingénierie et de gestion de Grenoble
Université Grenoble Alpes
46 avenue Félix Viallet - 38031 Grenoble - France
www.grenoble-inp.fr
© Copyright 2020 | Grenoble Institute of Technology (Grenoble INP) - Université de Grenoble, France.

Grenoble INP

UGA

Institut polytechnique de Grenoble
(Grenoble INP Institute of Engineering and Management)
46 avenue Félix Viallet

99% Unread: 2461 Total: 27698 29 Today Pane

Special domains: example

Reply

Reply All

Forward

Archive

Junk

Delete

More

From DGS | Grenoble INP <noreply@grenoble-inp.fr> ☆

Subject [SPAM HIGH ATTENTION]

To maciej.korczynski@grenoble-inp.fr ★

15:14

Avertissement de boîte aux lettres!

Vous avez atteint la capacité de stockage de votre courrier et ne pouvez plus recevoir de nouveaux messages tant que vous n'avez pas augmenté la capacité de stockage.

Veuillez suivre le lien ci-dessous pour augmenter la capacité de stockage.

[ACCROÎTRE LA CAPACITÉ](#)

Cordialement Votre,
Directrice Générale des Services
Chef de l'administration
Institut polytechnique de Grenoble (Grenoble INP)
Institut d'ingénierie et de gestion de Grenoble
Université Grenoble Alpes
46 avenue Félix Viallet - 38031 Grenoble - France
www.grenoble-inp.fr
© Copyright 2020 | Grenoble Institute of Technology (Grenoble INP) - Université de Grenoble, France.

Grenoble INP

UGA

Institut polytechnique de Grenoble

(Grenoble INP Institute of Engineering and Management)

46 avenue Félix Viallet

99%

Unread: 2461

Total: 27698

29 Today Pane

<https://webmail-grenoble-inp-fr.000webhostapp.com>

Special domains: example



Bienvenue sur Webmail Grenoble Institut de Technologie [Grenoble INP]

Entrez votre adresse e-mail et votre mot de passe pour vérifier votre compte

Adresse électronique :

Nom d'utilisateur :

Mot de passe :

Confirmez mot de passe :

Soumettre

Consequences of blocklisting URLs of special domain names

- When calculating security metrics (abuse rates) per registrars and TLDs, they must be excluded
- Decreased trust in intermediaries responsible for domain name registration

Special domain names

- Please use and contribute!
- Domain names are excluded from calculating abuse rates

korlabsio / subdomain_providers Public

<> Code Issues Pull requests Actions Projects Security Insights

main 1 branch 0 tags Go to file Code

File	Commit	Time
LICENSE	Initial commit	last year
README.md	Update IPFS domains	3 months ago
names	update	3 weeks ago

maroofi update b956f3c 3 weeks ago 61 commits

README.md

subdomain_providers

A curated list of free subdomain providers and dynamic DNS providers with their type

Description

This list provides active domain names that provide services to users like free hosting services (in terms of subdomains) or dynamic DNS providers as subdomain. The goal is to be able to detect a domain name as either user-registered domain or service provider.

The distinction is important since the take-down/mitigation action is different and can be performed by different entities.

For example, a domain name registered by an attacker to spread malware can be taken down by the registrar/registry. However, registries/registrar can not take down Google Drive free service in case of malware spread. Instead, one can notify Google for this case.

The list is provided in the [names](#) files in this repository. It's a two-column CSV file as follows:

1. first column: registered domain name
2. Second column: Type of service provider

korlabsio / urlshortener Public

<> Code Issues Pull requests Actions Projects Security Insights

main 1 branch 0 tags Go to file Code

File	Commit	Time
test	Initial Commit	last year
LICENSE	Initial commit	last year
README.md	Add commands to see duplicates	7 months ago
fetch.py	Initial Commit	last year
names.txt	update	last month

maroofi update db7e2b8 on Sep 18 30 commits

README.md

URLSHORTENER

A curated list of URL Shortener Websites with responses for crawling

Use Case

If you are writing a crawler or any kind of system that detecting a URL shortener is important, then this list can help you!

Description of the files

- names.txt

This file contains only the registered domain name for the URL shortener. The list **ONLY** contains the destination

False positives: Resilient Domain Whitelist

Building a Resilient Domain Whitelist to Enhance Phishing Blocklist Accuracy, J. Bayer, S. Maroofi, O. Hureau, A. Duda, M. Korczyński, to appear in eCRIME 2023

Principles:

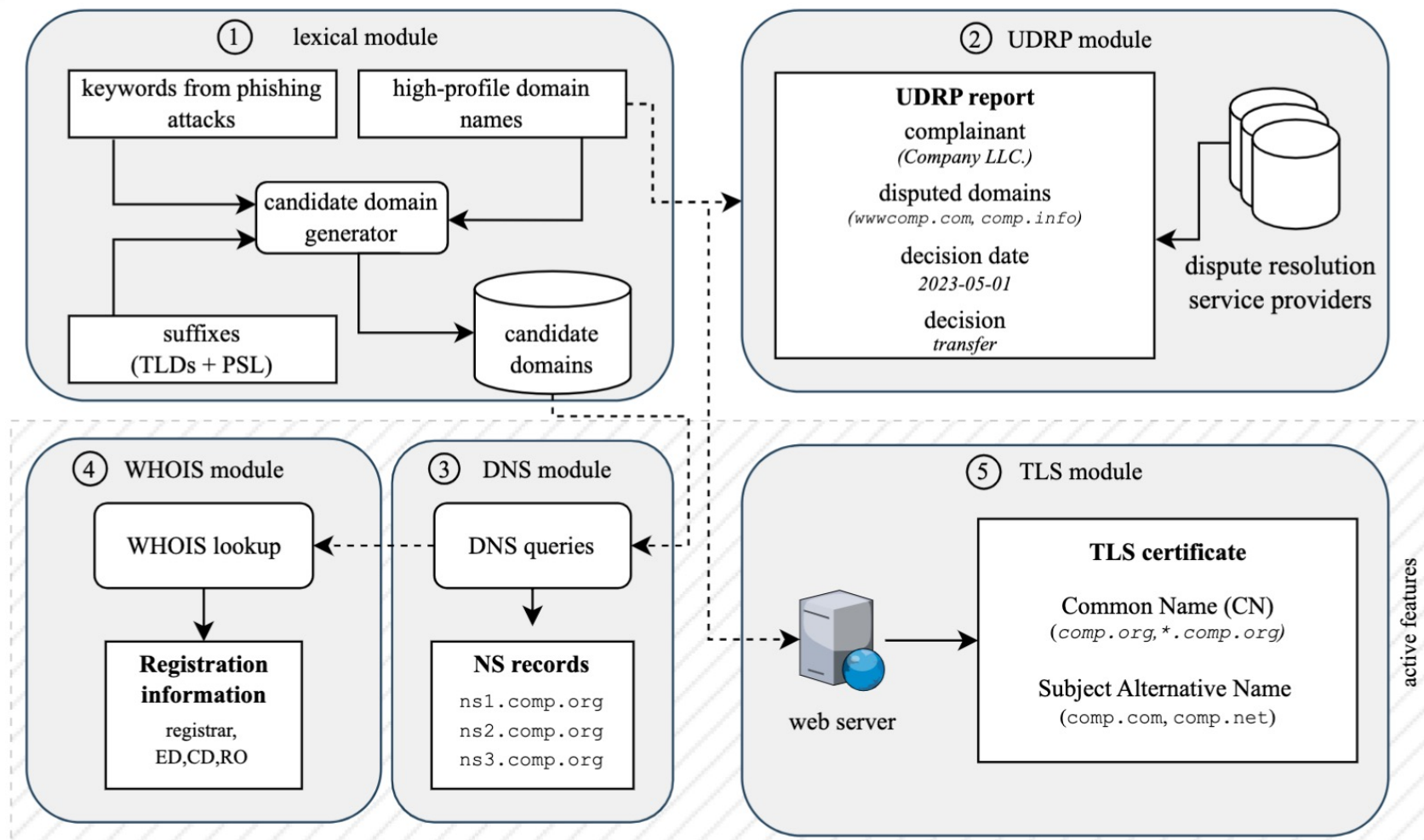
- Selection of curated seed domain names of high-profile organizations
- Procedure for validating disputed and defensively registered domain names

Datasets

Datasets used:

- Public DNS data
- Domain registration data (WHOIS)
- UDRP (Uniform Domain Name Dispute Resolution Policy) dispute-resolution service providers
- TLS (Transport Layer Security) certificates

Overview of the modules and data sources for building a domain whitelist



Lexical module: candidate domains

DS name	Description	Data source	Example	Size
$DS_{refDomains}$	high-profile domains targeted by phishing attacks	APWG [1], OpenPhish [2], PhishTank [3]	google.com ebay.com	338
$DS_{refBrands}$	brand names, i.e. e2LL	$DS_{refDomains}$	google ebay	305
DS_{kWords}	special keywords from domains used in phishing attacks	Bayer et al. [31]	support online secure	28
$DS_{candRoots}$	candidate roots generated by dnstwist (①)	$DS_{refBrands}$ [30]	góóglë ebay-help	1.69M
DS_{suffix}	active suffixes and TLDs	PSL [30]	org co.br	5,308
DS_{cand}	final candidates for squatting domains	$DS_{candRoots} \times DS_{suffix}$	gogle.org ebay.co.br	8.9B

Lexical module: candidate domains

DS name	Description	Data source	Example	Size
$DS_{refDomains}$	high-profile domains targeted by phishing attacks	APWG [1], OpenPhish [2], PhishTank [3]	google.com ebay.com	338
$DS_{refBrands}$	brand names, i.e. e2LL	$DS_{refDomains}$	google ebay	305
DS_{kWords}	special keywords from domains used in phishing attacks	Bayer et al. [31]	support online secure	28
$DS_{candRoots}$	candidate roots generated by dnstwist (①)	$DS_{refBrands}$ [30]	góóglë ebay-help	1.69M
DS_{suffix}	active suffixes and TLDs	PSL [30]	org co.br	5,308
DS_{cand}	final candidates for squatting domains	$DS_{candRoots} \times DS_{suffix}$	gogle.org ebay.co.br	8.9B

Lexical module: candidate domains

DS name	Description	Data source	Example	Size
$DS_{refDomains}$	high-profile domains targeted by phishing attacks	APWG [1], OpenPhish [2], PhishTank [3]	google.com ebay.com	338
$DS_{refBrands}$	brand names, i.e. e2LL	$DS_{refDomains}$	google ebay	305
DS_{kWords}	special keywords from domains used in phishing attacks	<i>Bayer et al.</i> [31]	support online secure	28
$DS_{candRoots}$	candidate roots generated by dnstwist (①)	$DS_{refBrands}$ [30]	góóglë ebay-help	1.69M
DS_{suffix}	active suffixes and TLDs	PSL [30]	org co.br	5,308
DS_{cand}	final candidates for squatting domains	$DS_{candRoots} \times DS_{suffix}$	gogle.org ebay.co.br	8.9B

Lexical module: candidate domains

DS name	Description	Data source	Example	Size
$DS_{refDomains}$	high-profile domains targeted by phishing attacks	APWG [1], OpenPhish [2], PhishTank [3]	google.com ebay.com	338
$DS_{refBrands}$	brand names, i.e. e2LL	$DS_{refDomains}$	google ebay	305
DS_{kWords}	special keywords from domains used in phishing attacks	<i>Bayer et al.</i> [31]	support online secure	28
$DS_{candRoots}$	candidate roots generated by dnstwist (①)	$DS_{refBrands}$ [30]	góóglë ebay-help	1.69M
DS_{suffix}	active suffixes and TLDs	PSL [30]	org co.br	5,308
DS_{cand}	final candidates for squatting domains	$DS_{candRoots} \times DS_{suffix}$	gogle.org ebay.co.br	8.9B

Lexical module: candidate domains

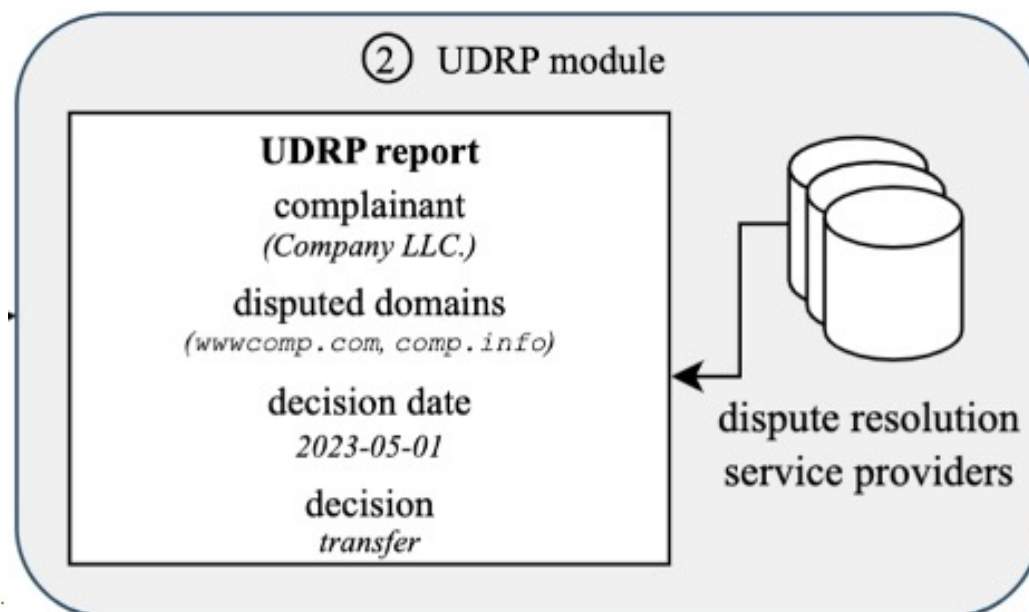
DS name	Description	Data source	Example	Size
$DS_{refDomains}$	high-profile domains targeted by phishing attacks	APWG [1], OpenPhish [2], PhishTank [3]	google.com ebay.com	338
$DS_{refBrands}$	brand names, i.e. e2LL	$DS_{refDomains}$	google ebay	305
DS_{kWords}	special keywords from domains used in phishing attacks	<i>Bayer et al.</i> [31]	support online secure	28
$DS_{candRoots}$	candidate roots generated by dnstwist (①)	$DS_{refBrands}$ [30]	góóglë ebay-help	1.69M
DS_{suffix}	active suffixes and TLDs	PSL [30]	org co.br	5,308
DS_{cand}	final candidates for squatting domains	$DS_{candRoots} \times DS_{suffix}$	gogle.org ebay.co.br	8.9B

Lexical module: candidate domains

DS name	Description	Data source	Example	Size
$DS_{refDomains}$	high-profile domains targeted by phishing attacks	APWG [1], OpenPhish [2], PhishTank [3]	google.com ebay.com	338
$DS_{refBrands}$	brand names, i.e. e2LL	$DS_{refDomains}$	google ebay	305
DS_{kWords}	special keywords from domains used in phishing attacks	<i>Bayer et al.</i> [31]	support online secure	28
$DS_{candRoots}$	candidate roots generated by dnstwist (①)	$DS_{refBrands}$ [30]	góóglë ebay-help	1.69M
DS_{suffix}	active suffixes and TLDs	PSL [30]	org co.br	5,308
DS_{cand}	final candidates for squatting domains	$DS_{candRoots} \times DS_{suffix}$	gogle.org ebay.co.br	8.9B

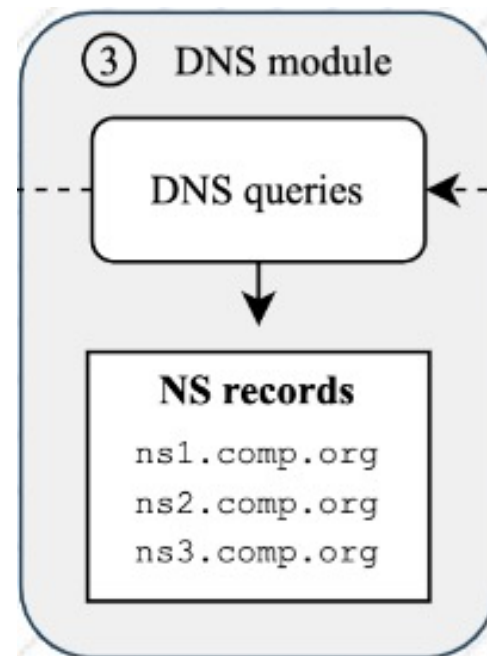
DD: Disputed Domain Names

- Dispute resolution (D-R) service providers:
 - ✓ World Intellectual Property Organization (WIPO)
 - ✓ Alternative Dispute Resolution (ADR)
 - ✓ Asian Domain Name Dispute Resolution Center (ADNDRC)
 - ✓ Canadian International Internet Dispute Resolution Center (CIIDRC)
- Extracted 135,043 unique disputed domains
- Whitelisted **7,925** domain names using the DD approach



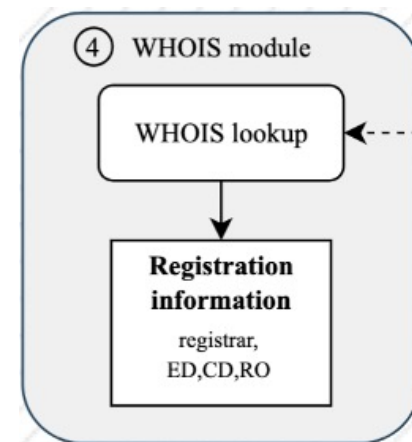
SINS: Shared In-Bailiwick Name Servers

- Collection of the NS records for the 8.9B candidate domains (DS_{cand})
 - ✓ 522.4K returned NOERROR, and a non-empty answer
- Collection of the NS records for the 338 reference domains ($DS_{refDomains}$)
- Reference domains whose name servers are in-bailiwick and in-domain, for example:
 - ✓ `google.com` is considered as in-bailiwick as its name servers `ns[1-4].google.com` are subdomains of the origin `google.com`
 - ✓ NS records for `visa.net` are `ns[1-3].dnsvisa.com` indicating out-of-bailiwick name servers
- Whitelisted **5,209** domain names using the SINS approach



DR: Defensive Registrars

- For each of the 2.5 million registered domains (among 8.9B candidate domains), we have retrieved the Registrant Organization (RO) name (from WHOIS)
- If the RO of the candidate domain matches that of the reference domain names, and the Registrar is one of the 9 defensive registrars, the domain is whitelisted
- Whitelisted **3,979** domain names using the SINS approach

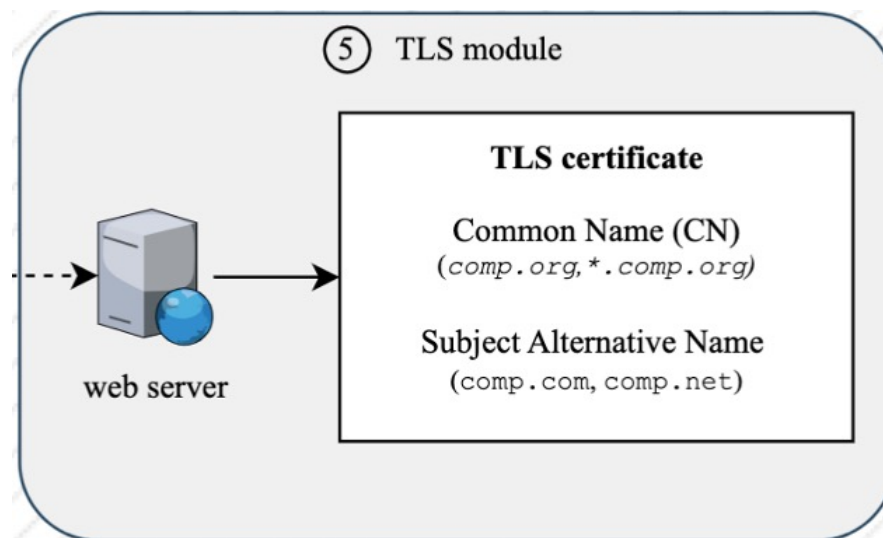


Registrar name	IANA ID
MarkMonitor	292
CSC Corporate Domains	299
Com Laude	470
RegistrarSEC	2475
Safenames	447
Nameshield	1251
IP Twins	1728
SafeBrands	1290
Hogan Lovells	1526

Defensive registrars

DC: Domain Certificates

- Two options for specifying the domain name or a wildcard: Common Name (CN) field of the Subject or `subjectAltName` extension
- For the 338 reference domains we collected a dataset of 6,946 entries, including 6,071 fully qualified domain names (FQDNs)
- Cross-reference this list with the domains listed in candidate domains (DS_{cand})
- Whitelisted **129** domain names using the DC approach



Whitelist composition: further notes

- Domain validity period
- Types of entries:
 - ✓ fully qualified domain names
 - ✓ wildcards

Evaluation

- 17,215 domains in the final whitelist created on September 1, 2023
- Disputed domains (DD) method contribute the most to the final whitelist.
- The domains generated for *Amazon, Google, Microsoft, Instagram, and Facebook* together account for 32.3% of the whole whitelist
- Confirmed false positives in existing blacklists: examples

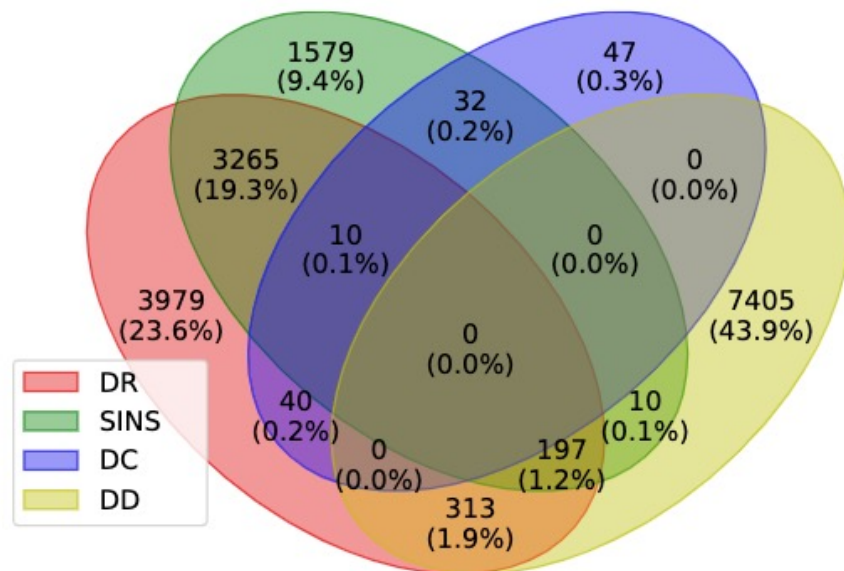


Figure 2. Overlap of whitelisted domains coming from different sources

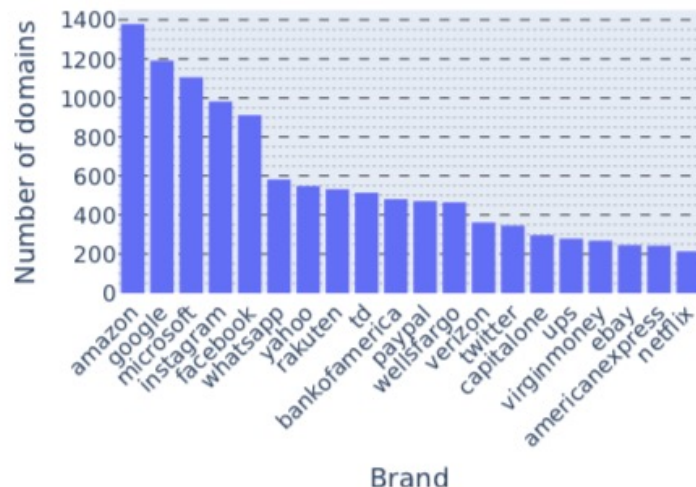


Figure 3. Top 20 brands with the most important contribution to the whitelist

Summary

- Problem of false positives, phishing simulations, and special domain names on blocklists
- Proposed approached to limit their consequences
- Building a Resilient Domain Whitelist to Enhance Phishing Blocklist Accuracy, J. Bayer, S. Maroofi, O. Hureau, A. Duda, M. Korczyński, to appear in eCRIME 2023

Questions?

maciej.korczynski@univ-grenoble-alpes.fr