Building a Resilient Domain Whitelist to Enhance Phishing Blocklist Accuracy

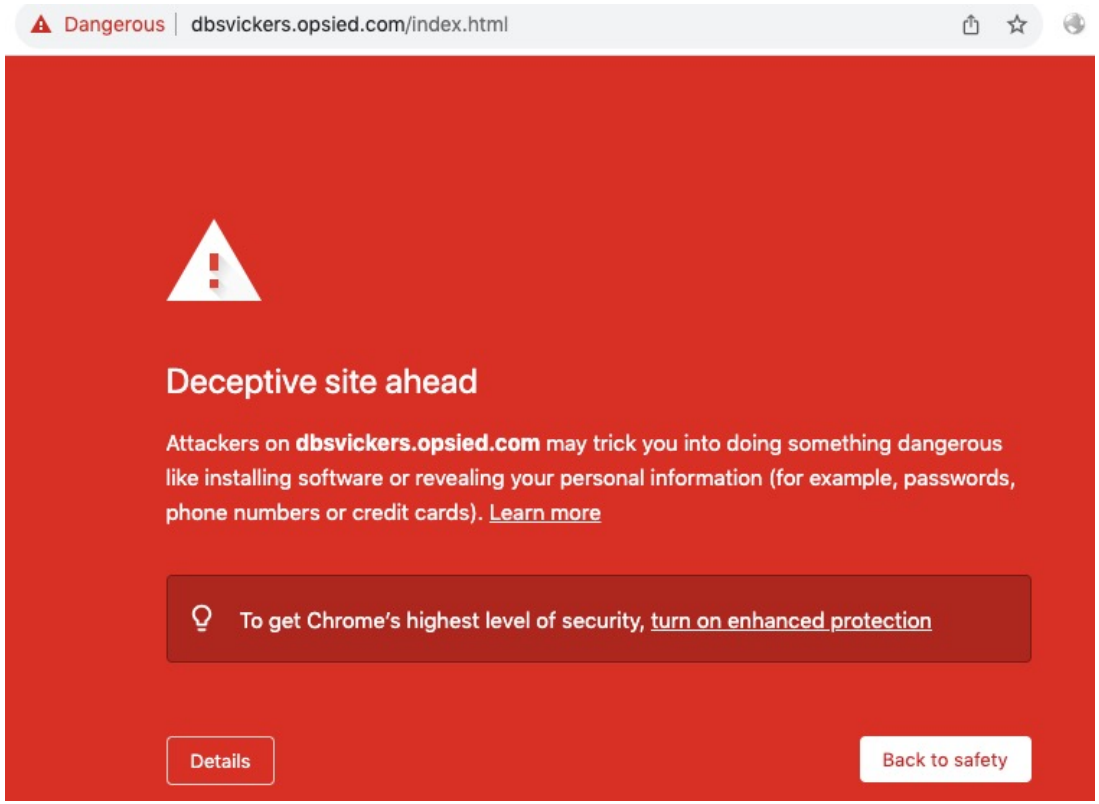
Maciej Korczyński, Sourena Maroofi, Jan Bayer, Andrzej Duda

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

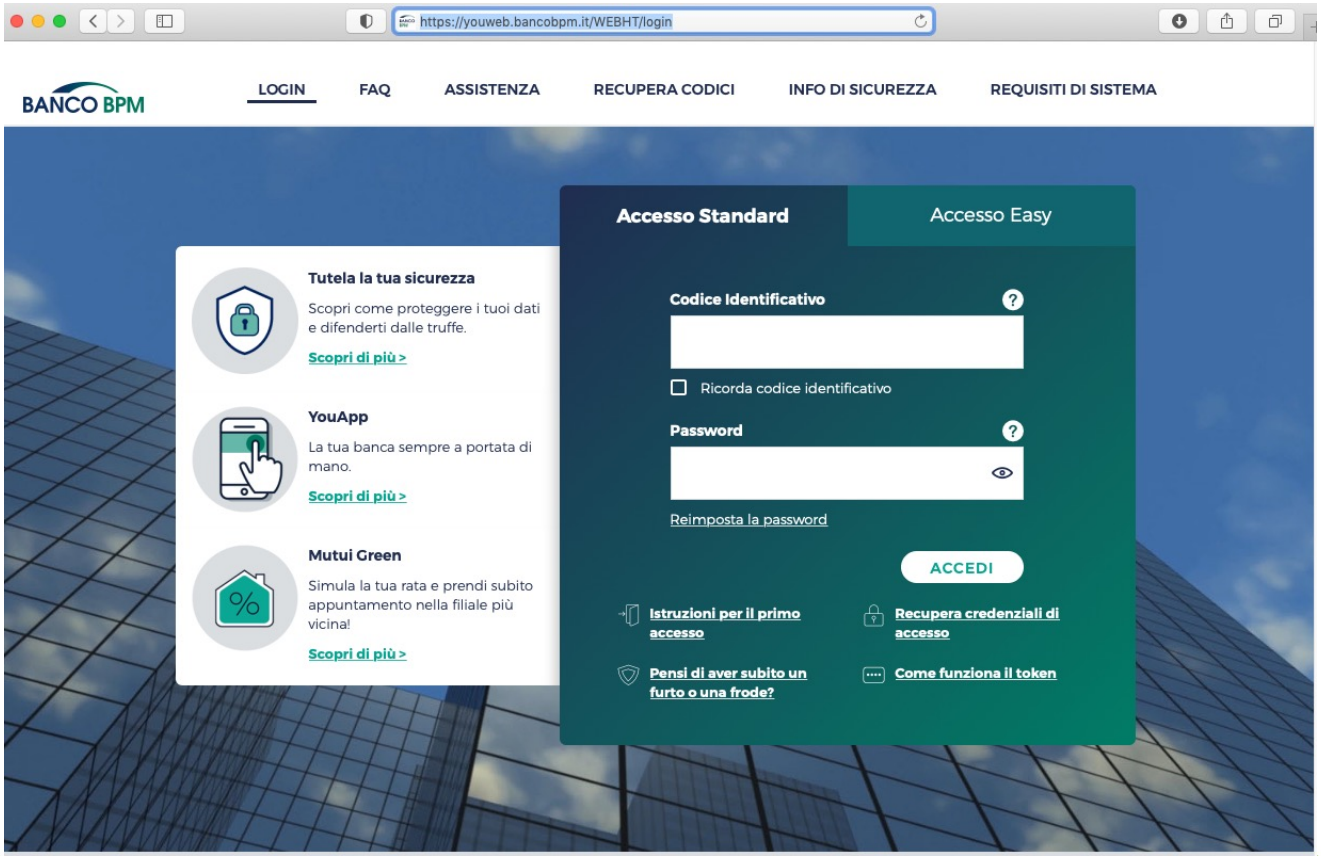
Agenda

- 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

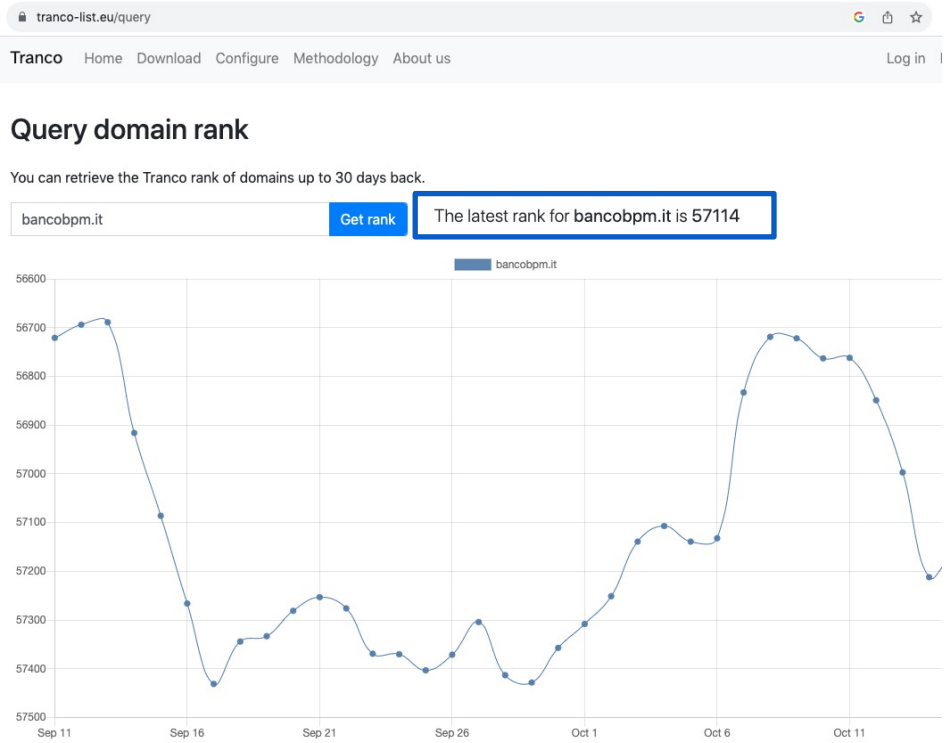
- URLs and domain names incorrectly categorized as malicious

```
whois bancobpm.it
Created:    2016-03-28
```

False positives: example

- URLs and domain names incorrectly categorized as malicious

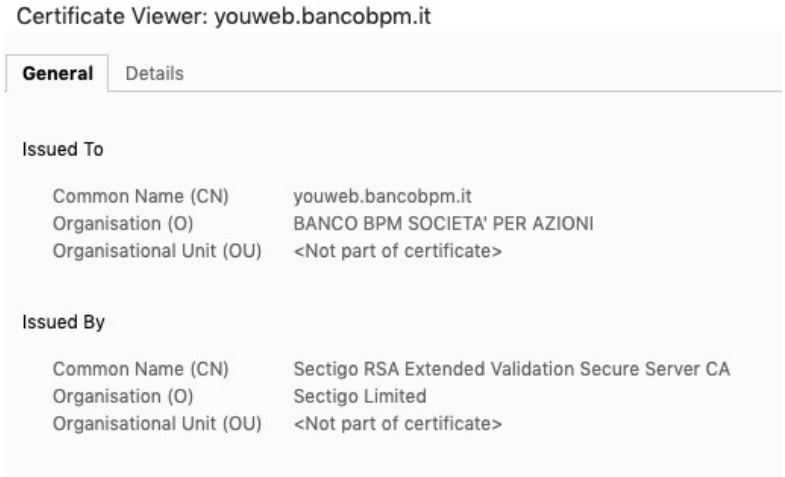
```
whois bancobpm.it
Created:    2016-03-28
```

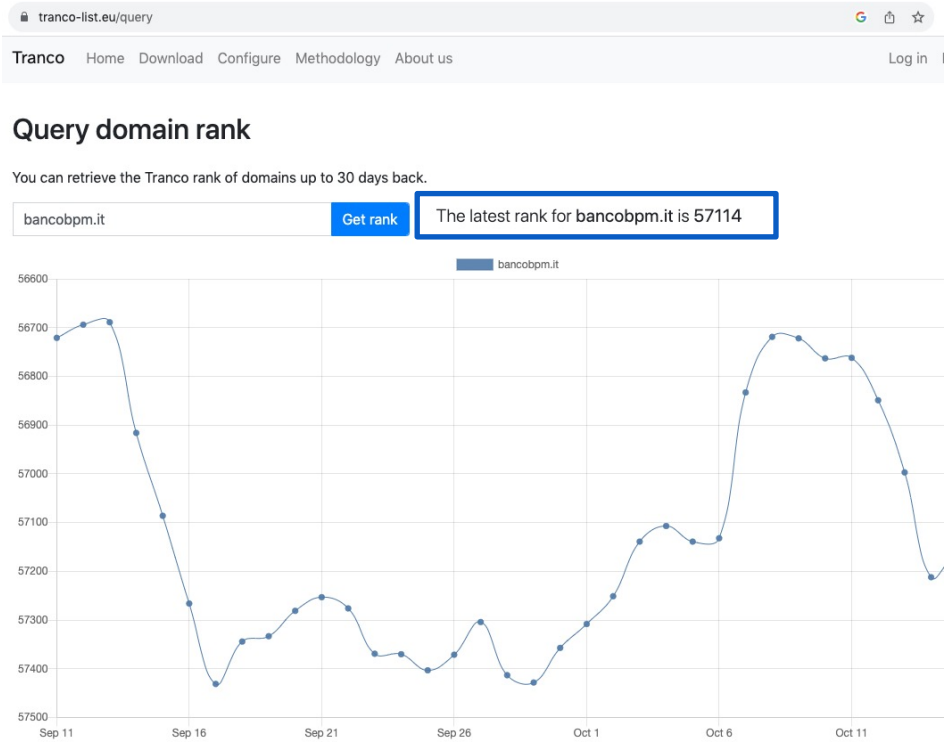


False positives: example

- URLs and domain names incorrectly categorized as malicious

```
whois bancobpm.it
Created:    2016-03-28
```





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

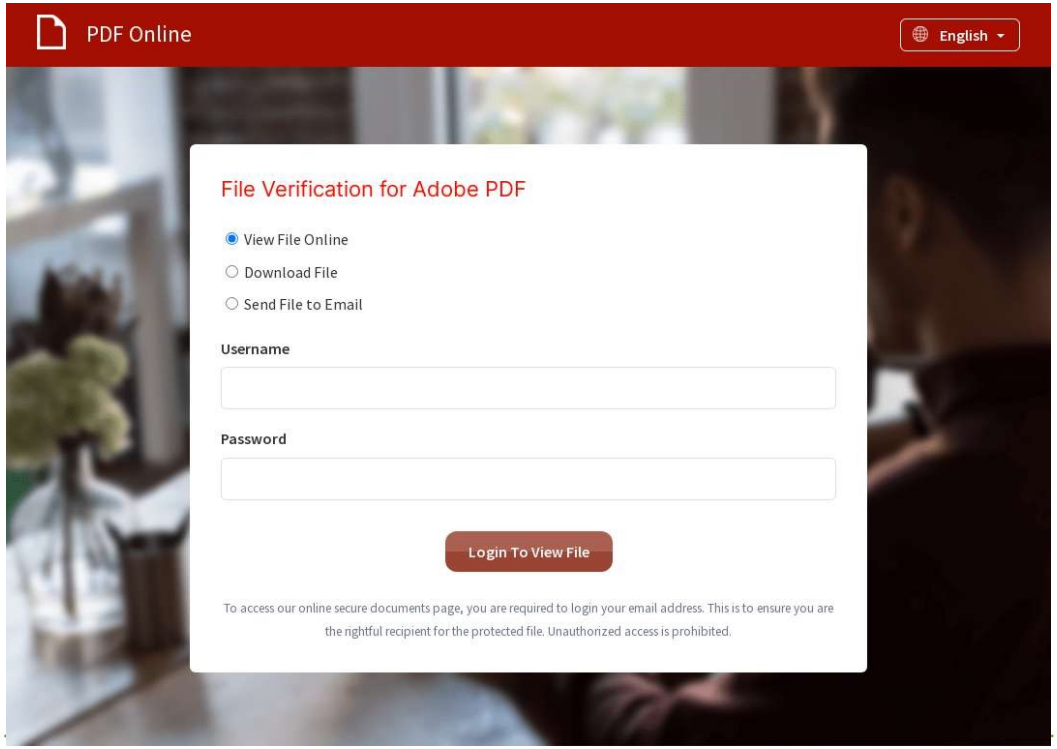
Consequences of false positives

- 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



Phishing simulation: example

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

Blockilisting 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

Blockilisting 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

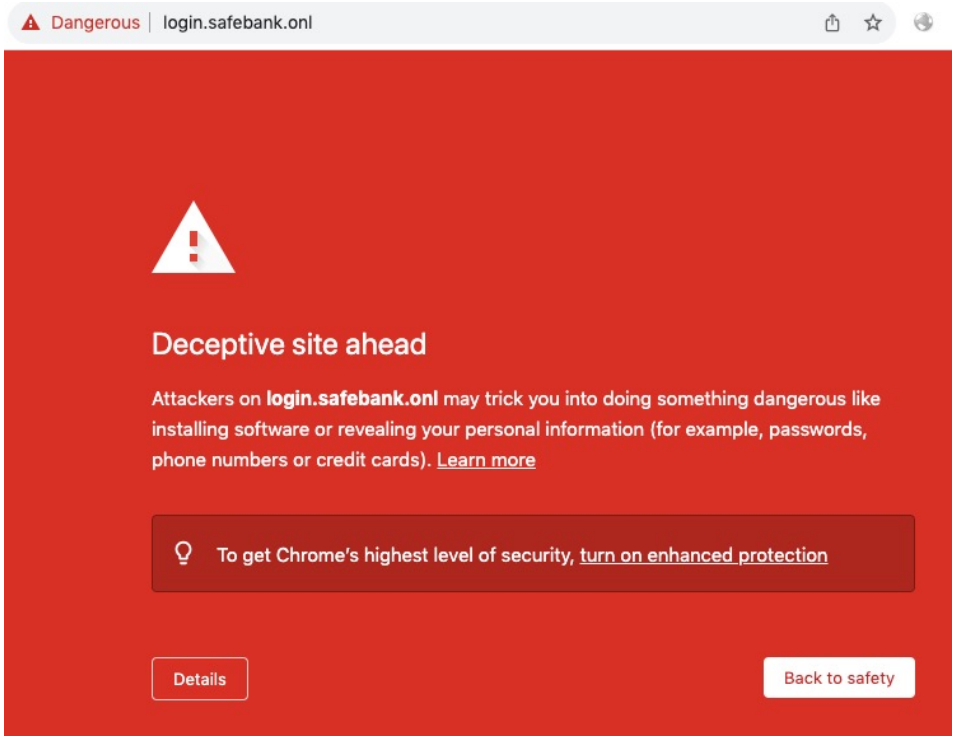
Blockilisting date: 2023-07-29

```
whois safebank.onl
Creation Date: 2015-07-17
Registrar: MarkMonitor Inc.
Registrar IANA ID: 292
```

Web content of login.safebank.onl

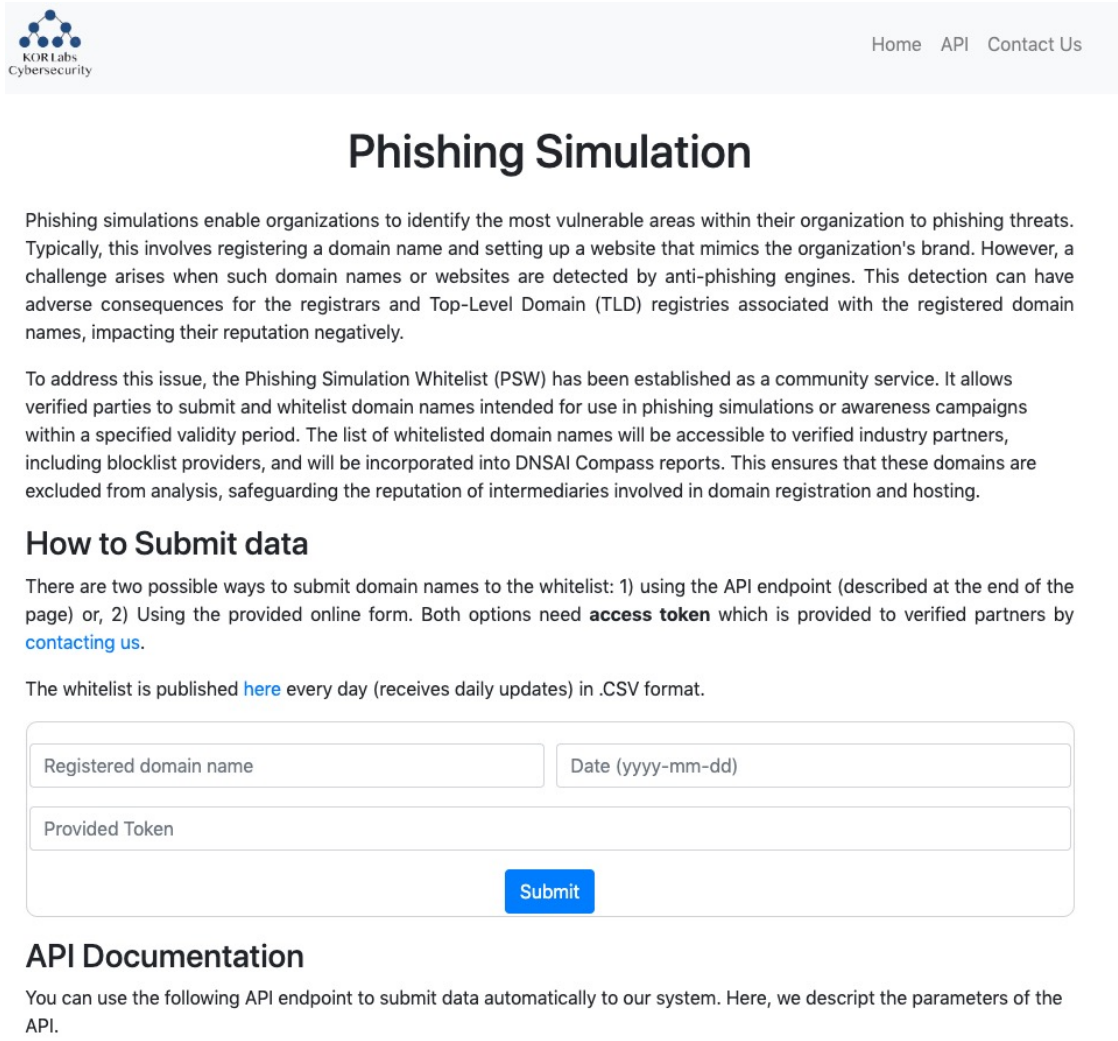
Consequences of blocklisting phishing simulation websites and domain names

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



Phishing simulation: community feed

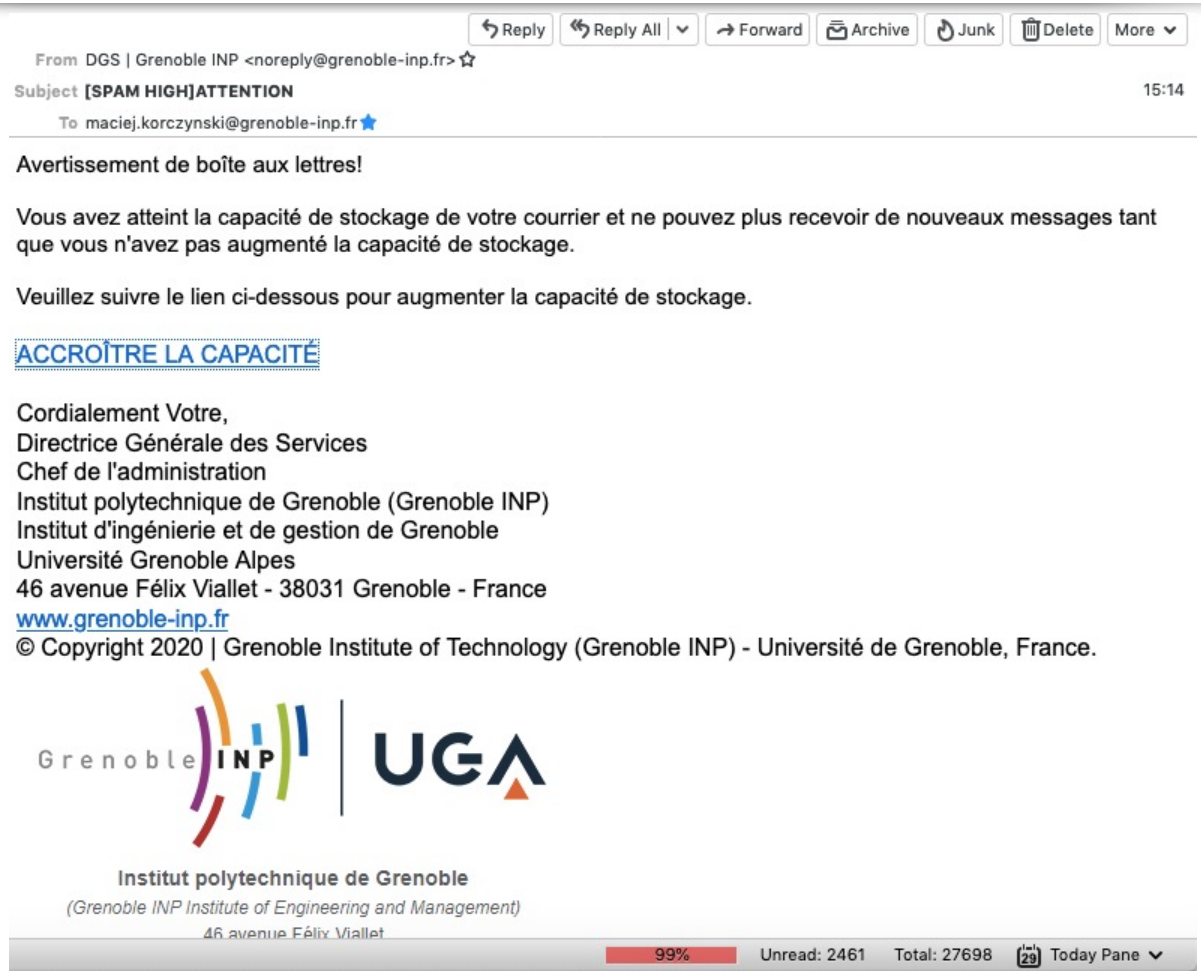
- Community service: Phishing Simulation Whitelist (PSW) allows verified parties to submit and whitelist domain names intended for use in phishing simulations or awareness campaigns within a specified validity period
- Usage: PSW accessible by everyone, will be used in DNSAI COMPASS



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 INP | UGA

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>

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.

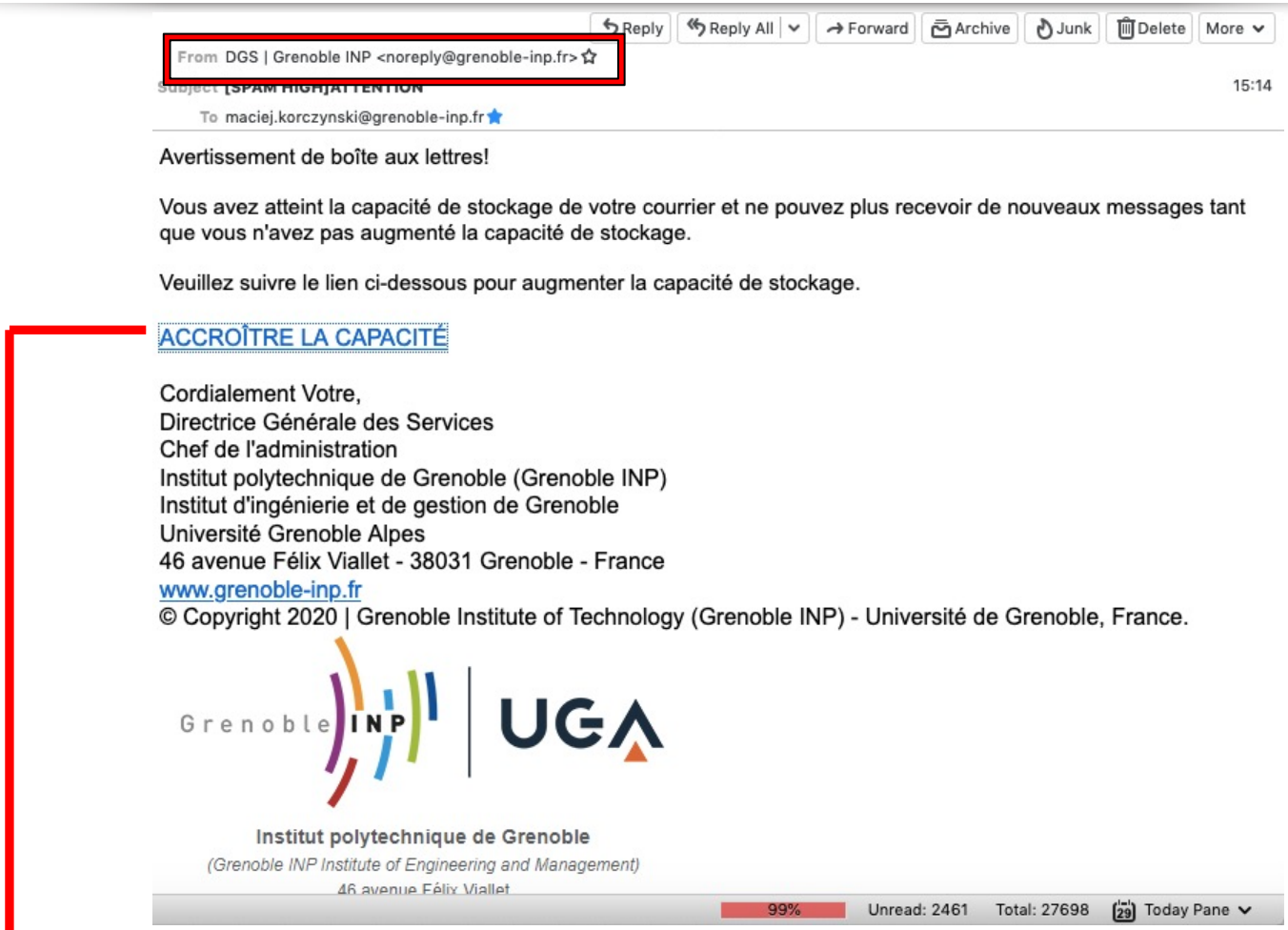
ACCROÎTRE LA CAPACITÉ

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.

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



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

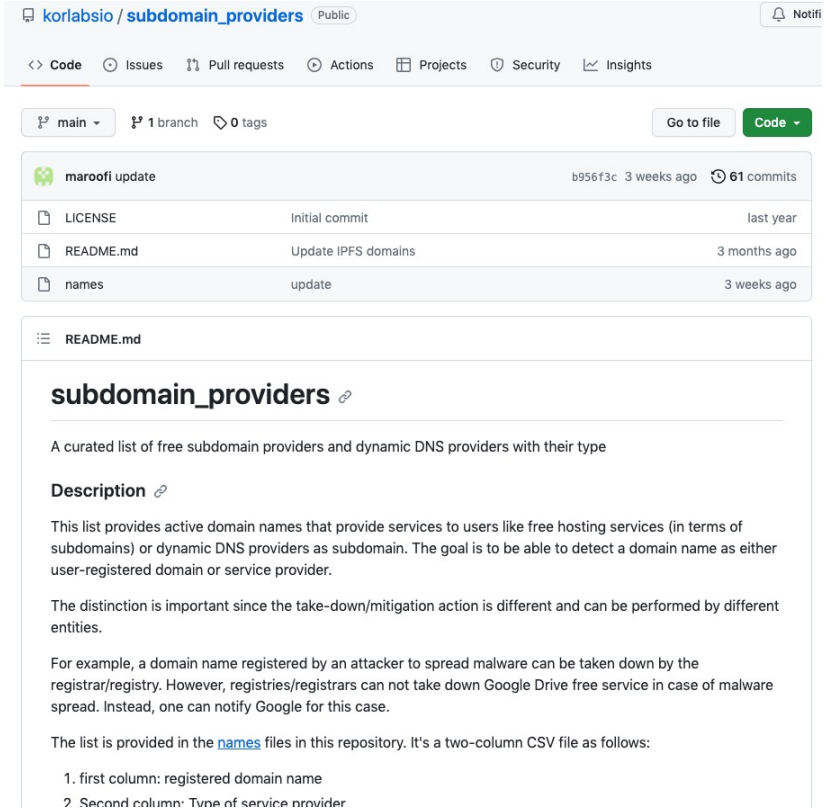
Special domains: example

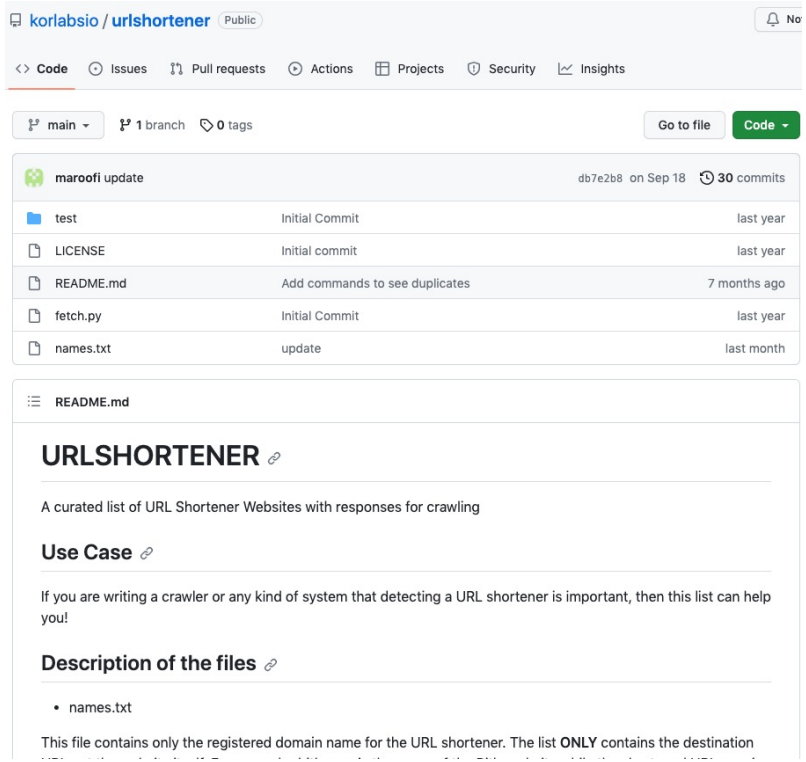
Consequences of blocklisting URLs of special domain names

- 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





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:

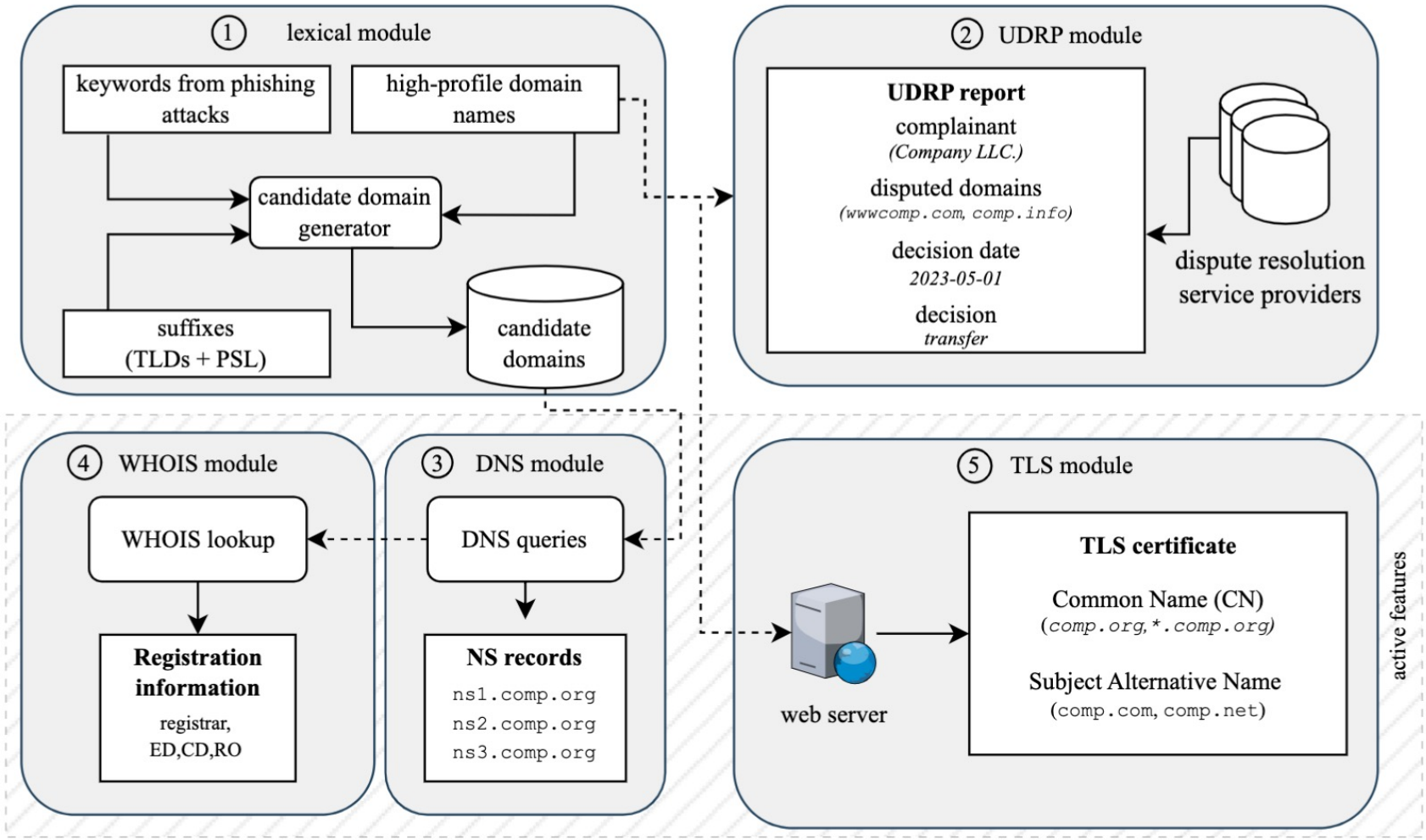
- Selection of curated seed domain names of high-profile organizations
- Procedure for validating disputed and defensively registered domain names

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	*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

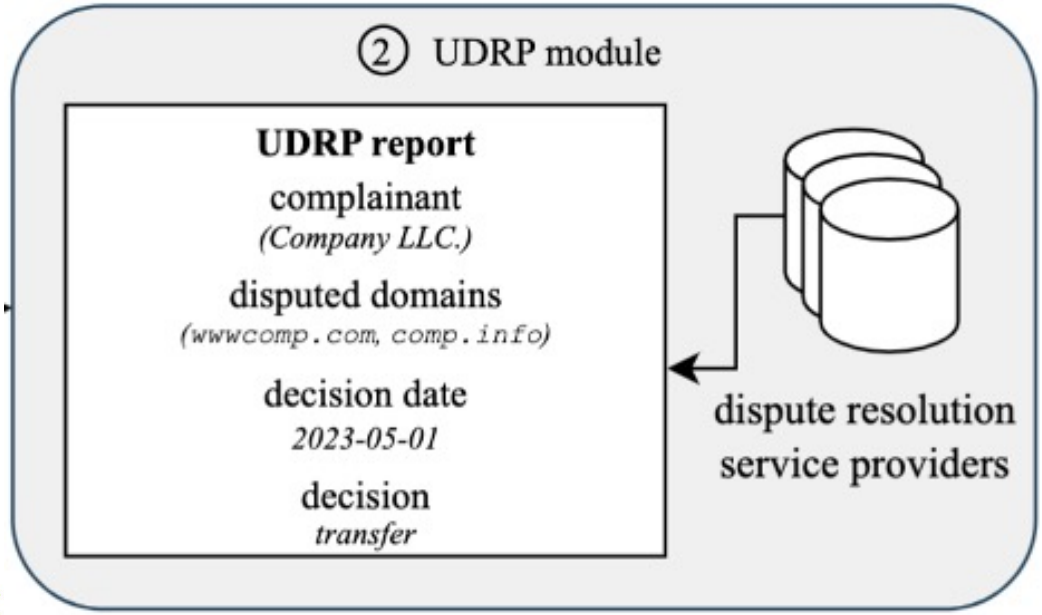
DS name	Description	Data source	Example	Size
$DS_{refDomains}$	high-profile domains targeted by phishing attacks	APWG [1], OpenPhish [2], PhishTank [3]	`google.com` `ebay.com`	338
$DS_{refBrands}$	brand names, i.e. e2LL	$DS_{refDomains}$	`google` `ebay`	305
DS_{kWords}	special keywords from domains used in phishing attacks	*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

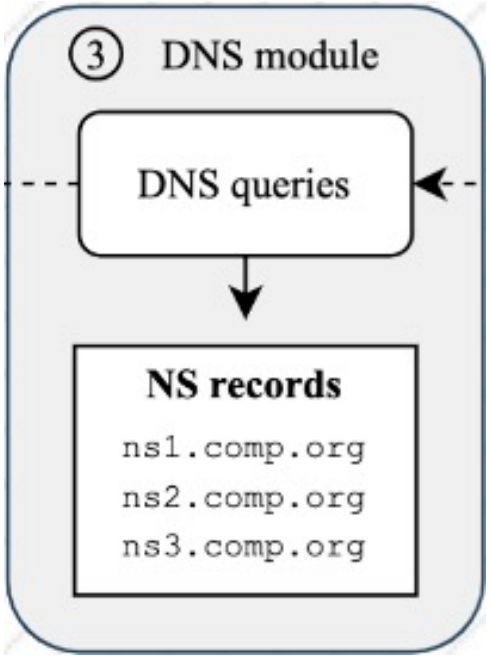
DD: Disputed Domain Names

- 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



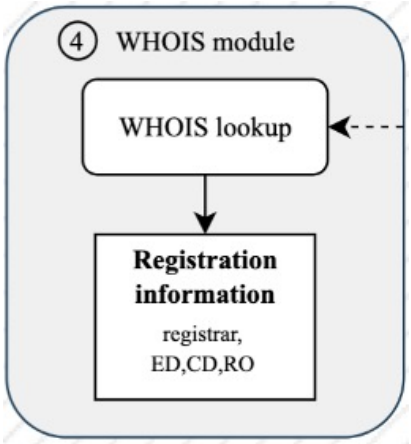
SINS: Shared In-Bailliwick Name Servers

- 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-baliwick 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

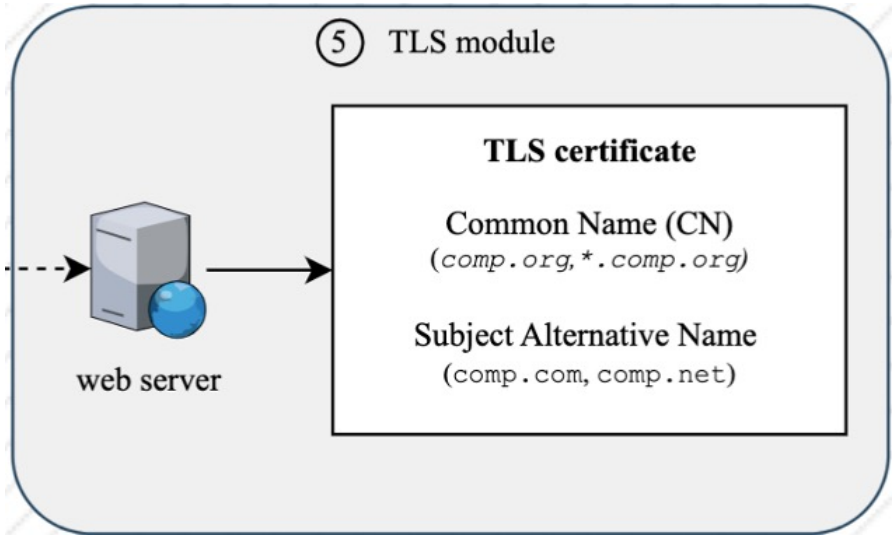


Registrar name	IANA ID
MarkMonitor	292
CSC Corporate Domains	299
Com Laude	470
RegistrarSEC	2475
Safenames	447
Nameshield	1251
IP Twins	1728
SafeBrands	1290
Hogan Lovells	1526

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

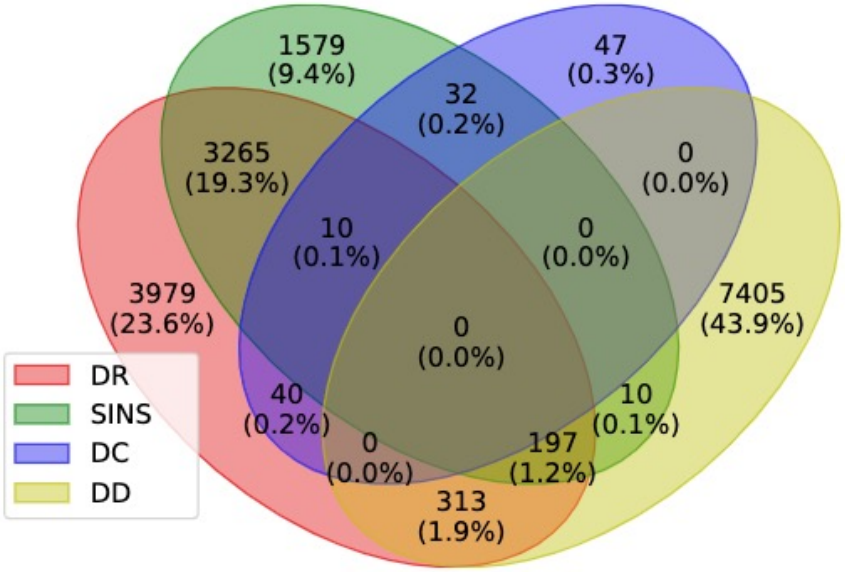


Figure 2. Overlap of whitelisted domains coming from different sources

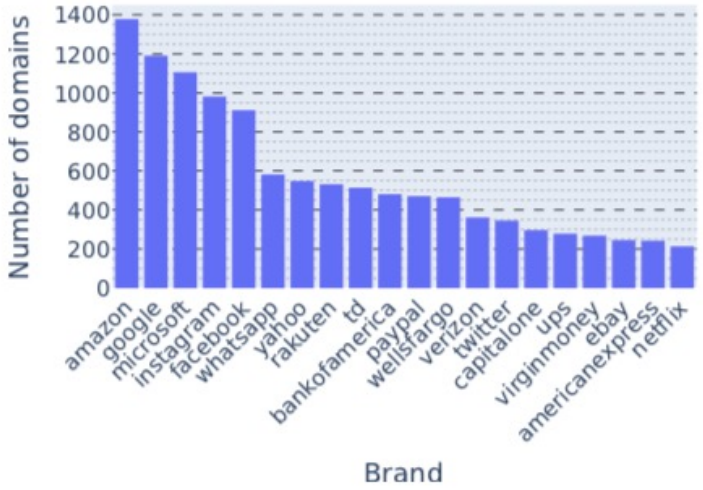


Figure 3. Top 20 brands with the most important contribution to the whitelist

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

Grenoble INP ENSIMAG

Questions?

maciej.korczynski@univ-grenoble-alpes.fr