

# Typosquatting Finder

An Open Source Solution to Find Typosquatted Domains



**CIRCL**

Computer Incident  
Response Center  
Luxembourg

Team CIRCL  
*TLP:CLEAR*

[info@circl.lu](mailto:info@circl.lu)

20230704

## History - initial use-cases

---

- AIL project<sup>1</sup> is composed of **feeders acting as source for further collection and analysis** in AIL framework
- A use case with the AIL certificate transparency feeder<sup>2</sup> was originally developed
- Detect typo-squatted domains<sup>3</sup> generated from known constituents is used for filtering and detection in AIL

---

<sup>1</sup><https://www.ail-project.org/>

<sup>2</sup><https://github.com/ail-project/ail-feeder-ct>

<sup>3</sup><https://github.com/typosquatter/ail-typo-squatting>

## Why typosquatting finder was developed?

---

- **No common open source library** for all the known or less-known typosquatting techniques
- No easy way to check and **proactively automate potential typosquatted domains** into a CTI pipeline
- Integrating typosquatting cases in other tools (e.g. finding typosquatted package names<sup>4</sup> in npmjs, pypi)

---

<sup>4</sup> *most malicious code gets into applications through library "typosquatting"*

<https://arxiv.org/abs/2005.09535>

## Components designed and developed

---

- A python library (ail-typo-squatting) to include **all known typosquatting algorithms**
- A **website to find potential typo-squatted domains**<sup>5</sup> with export as MISP feed or MISP JSON
- Analyse discovered domains for **similarities with the original website**
- A tool to find typosquatted packages in PyPI<sup>6</sup>
- An **extension to AIL** to include all the domain variations

---

<sup>5</sup><https://typosquatting-finder.circl.lu/>

<sup>6</sup><https://github.com/typosquatter>

## 21 known typosquatting algorithms implemented

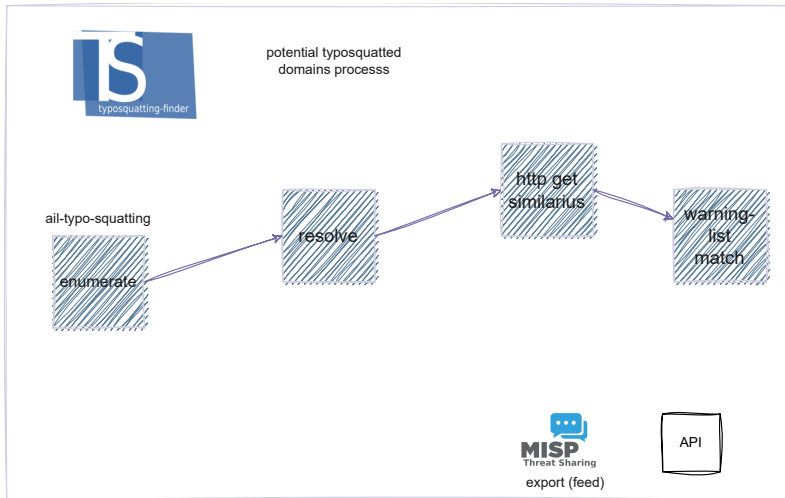
---

- AddDash, Addition, AddDynamicDns, AddTld, ChangeDotDash, ChangeOrder, CommonMisspelling, Double Replacement, Homoglyph, Homophones
- MissingDot, NumeralSwap, Omission, Repetition, Replacement, StripDash, SingularPluralize, Subdomain, VowelSwap, WrongTld, WrongSld
- Missing one? don't hesitate to **propose your own algorithm or open a pull-request**<sup>7</sup>.

---

<sup>7</sup><https://github.com/typosquatter/ail-typo-squatting>

# Typosquatting - ail-typo-website overview



ail-typo-website

## Similarius - comparing web pages

---

- The typosquatting finder has the capability to uncover numerous potential typosquatted domains
- However, analysts would need **to validate these findings**, which can be a time-consuming process
- Similarius<sup>8</sup> was developed to compare web pages and evaluate their level of similarity
- The similarity assessment (ratio) relies on two techniques:
  - Comparing the matrix of token counts between the original website and the discovered one
  - Comparing the resources (hyperlinks) between the two websites







---

<sup>8</sup><https://github.com/ail-project/Similarius>

# Parking pages removal with MISP warning-lists

---

- Typosquatted domains are not directly used for malicious activities
- Registration can occur months before any actual abuse takes place
- MISP warning-lists<sup>9</sup> have been expanded to identify parking page NS records and IP infrastructure

 kbc.ai wrongTld	  86.105.245.69 Netherlands	ns2.eftydns.com <a href="#">1 more...</a>	KBC.ai domain name is for sale. Inquire now.	6 %	100 %	0.36 %
 kbc.com   wrongTld	2.17.197.162 <a href="#">more...</a> Netherlands	a9-66.akam.net <a href="#">5 more...</a>	5 mail61.mailinfra. com <a href="#">5 more...</a>	16 %	91 %	2.81 %

---

<sup>9</sup><https://github.com/MISP/misp-warninglists>



# Typosquatting finder website - query interface

- CIRCL operates a publicly accessible typosquatting finder website<sup>10</sup>, which can also be easily installed on-premises<sup>11</sup>



mybank.com Search

Found 6128 domains, identified 133 registered.

Advanced Share Download +

- Run all algorithm
- Add Dash
- Addition
- BHSquatting
- Change Order
- Double Replacement
- Homophones
- Missing Dot
- Repetition
- Singulare Pluralize
- Subdomain
- Vowel Swap
- Add TLD
- Add dynamic DNS
- Change Dot to Dash
- Common Misspelling
- Homoglyph
- Keyboard Insertion
- Omission
- Replacement
- Strip Dash
- Transposition
- Wrong TLD

Enter list of NS and/or list of MX separate by comma. Domains will be mark in case of match.

NS ns1.exemple.com

MX mail.exemple.com

[\[more info\]](#)

<sup>10</sup><https://typosquatting-finder.circl.lu/>

<sup>11</sup><https://github.com/typosquatter/ail-typo-website>





# Typosquatting finder website - results

bank-of-america.com 100% Search

Found 6444 domains. Identified 458 registered.

Advanced Share Download ▾



- List of Variations
- Domain Identified
- Misp Feed
- Misp Json

PERMUTATION	IP ADDRESS	NAME SERVER	MAIL SERVER	WEB TITLE	WEB SIMILARITY	RESSOURCE DIFF
bank-of-america.com   original		dns2.cscdns.net <a href="#">1 more...</a>				
bnk-of-america.com  		dns2.cscdns.net <a href="#">1 more...</a>				

# Typosquatting finder website - example

---

...eng...

bank-of-america.appspot.com	142.250.219.52 2800:3f0:4001:82c::20	5 gmr-smtp-in.l.google.com	Login
 	14 United States	<a href="#">4 more...</a>	
wrongTld			

**Figure:** Typosquatting finder result



**Figure:** Website<sup>12</sup>

---

<sup>12</sup><https://lookyloo.circl.lu/tree/095d76be-6067-4d39-bac2-3f967f4fb54d>

## Package repository typosquatting - PyPI

---

- Typosquatting of package or module names has become a recurring problem<sup>13</sup>
- **Applying typosquatting techniques** to package names is somehow similar to domain typosquatting
- We developed a PyPI-squatting tool<sup>14</sup> to efficiently search for typosquatted modules

---

<sup>13</sup>[https:](https://blog.phylum.io/a-pypi-typosquatting-campaign-post-mortem/)

[//blog.phylum.io/a-pypi-typosquatting-campaign-post-mortem/](https://blog.phylum.io/a-pypi-typosquatting-campaign-post-mortem/)

<sup>14</sup><https://github.com/typosquatter/pypi-squatting>

## Typographic errors in other sources

---

- The typosquatting finder can be used to identify **mistyped domains or ongoing campaigns** in forums or other sources.
- The typosquatting finder library is integrated into the AIL project<sup>15</sup> as a tracking tool.

---

<sup>15</sup><https://github.com/ail-project/ail-framework>

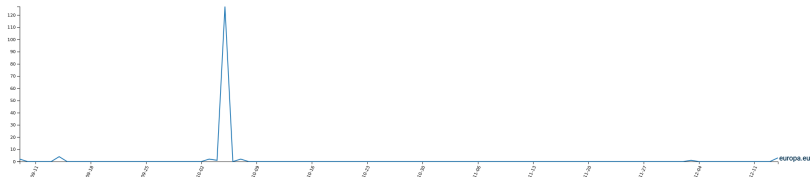
# All project integration

europa.eu typo [✎](#)  
fdbx22d3152-4691-9c00-8cc01d3f268d

Type	Tracker	Date added	Access Level	Created by	First seen	Last seen	Tags <a href="#">✎</a>	Email <a href="#">✎</a>
typosquatting	<a href="#">europa.eu</a>	2022/09/09	Global	aurelien@circl.lu	2022/09/09	2022/12/14	<a href="#">topic=europa</a>	

Sources:  
[All sources](#)

[Edit Tracker](#) [✎](#) [🔴](#)



20220909

[Search Tracked Items](#)

Show 10 entries

Search:

Date	Item id
2022/09/09	submitted/2022/09/09/submitted_46355bed-88a8-4f2f-9b9d-dbd998f33fd.gz <a href="#">topic=europa</a> <a href="#">title=europa.eu</a>
2022/09/09	submitted/2022/09/09/submitted_a9ca51c6-19b8-4d9a-9469-3b13826e44b.gz <a href="#">title=europa.eu</a> <a href="#">url=https://www.europa.eu</a>

# AIIL project integration

The screenshot displays the AIIL project interface. At the top, there is a navigation bar with links for Home, Submit, Tags, Links Hunter, Crawlers, Objects, Server Management, and Log Out. A search bar is located on the right. Below the navigation bar, a dark header shows the file path: `crawled / 2023 / 06 / 04 / leaksndi6i6m2j6ozulqe4imlrq6wrgjlhxe25vremvr3aymm4aaid.onion005babba-1d46-425a-9701-2e9e87f2ae6f`.

The main content area features a table with the following columns: Date, Source, Size (Kb), Number of lines, and Max line length. The table contains one entry:

Date	Source	Size (Kb)	Number of lines	Max line length
2023/06/04	crawled	16.19	592	419

Below the table, there are two buttons: `urlleak:automatasy-detectipn="ipart"` and `urlleak:submitssom="crawler"`. A "Father" link is provided: `crawled/2023/06/03/catalogwefcc5ngp3m3ngjps3dcknuf57xsaawf8epw3ymqf.onion012970a-101d-4731-845d-6dbb165864`. Action buttons include "Add to Export", "Investigations", and "Correlations Graph".

There is a "duplicates" section with a toggle and a "Crawler" section with a toggle. The "Last Origin" section lists:

- `leak64782376153vatabgvevufcay3m3ngjps3dcknuf57xsaawf8epw3ymqf.onion012970a-101d-4731-845d-6dbb165864`
- `leak6465662646adpgebdagfay3m3ngjps3dcknuf57xsaawf8epw3ymqf.onion`

The URL `url: http://leakndi6i6m2j6ozulqe4imlrq6wrgjlhxe25vremvr3aymm4aaid.onion/` is also shown.

The "Hacked databases store" section includes a "Full resolution" button and a table of hacked databases:

Year	Database	Site	Records	Price	Buy
2022	Shopee users and credit cards <b>NEW!</b>	shopee.com	257,026,348	\$200	Buy
2022	Twitter January 2022 leak <b>NEW!</b>	twitter.com	226,047,943	\$278	Buy
2022	00Webroot Database	00webroot.com	13,540,468	\$48	Buy
2022	123RF Database	123rf.com	6,963,378	\$45	Buy
2021	123RF Credit Database	123rf-credit.com	28,965	\$20	Buy
2019	1433978 Database	1433978.com	198	\$19	Buy
2018	11 Media (1) DB Database	11-media.com	28,052,122	\$82	Buy
2015	171710 Chrome Database	17171.com	5,792,608	\$42	Buy
2015	176 Database	176.com	6,070,071	\$41	Buy

## Conclusion

---

- **Automatically discovering newly registered typosquatted** domains and pushing them into MISP for evaluation is highly beneficial
- The **library currently outperforms**<sup>16</sup> other tools in its domain
- In addition, it provides an extra advantage by offering access to an **additional Passive DNS source**
- A valuable lesson learned from this experience is the advantage of initially developing a library and then integrating it into your toolset

---

<sup>16</sup><https://github.com/TiiTcHY/TypoSquat-Domain-Comparison>



## References

---

- Online public service  
<https://typosquatting-finder.circl.lu/>
- Source code of the library and services  
<https://github.com/typosquatter>
- Contact: [info@circl.lu](mailto:info@circl.lu)
- Thanks to David Cruciani for the endless search of typosquatting techniques. Thanks to European Commission Directorate General for Informatics for the initial challenge!
- The project has been co-funded by CEF-TC-2020-2 - 2020-EU-IA-0260 - JTAN - Joint Threat Analysis Network