AIL Project

Practical and Efficient Data-Mining of Chats, Suspicious Websites, Forums and Tor Hidden-Services

CIRCL
Computer Incident
Response Center
Luxembourg

**Co-funded by
the European Union**

Team CIRCL
info@circl.lu

CIRCL

July 10, 2024

Background

- Over the past years, CIRCL has developed the AIL project[1] to fulfill our needs at CIRCL in intelligence gathering and analysis.
- AIL features an extensible Python-based framework for the **analysis of unstructured information**, collected either through an advanced crawler manager or from various feeders, including social networks and custom feeders.
- The AIL Project is an **open-source** framework[2] comprising various modules designed for the **collection, crawling, digging, and analysis of unstructured data**.
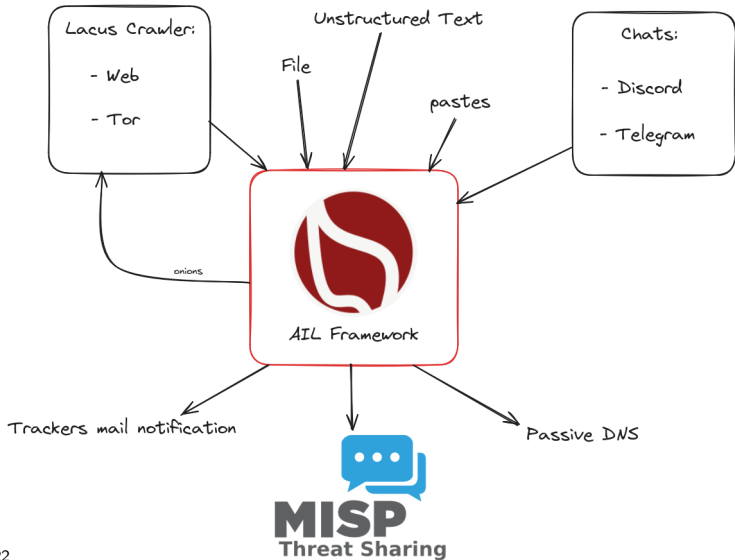

ail project

---

[1] https://www.ail-project.org/
[2] https://github.com/ail-project

High Level Overview

Collection - automate crawling

- Crawling can be a challenging task, for example, gathering all the blog posts from ransomware groups[3], which can be demanding for an analyst.
- AIL offers a crawling feature that can **initiate regular crawls using a standard spawned browser**.

---

[3]https://www.ransomlook.io/

Collection - Lacus Crawler[4]

---

[4]https://github.com/ail-project/lacus
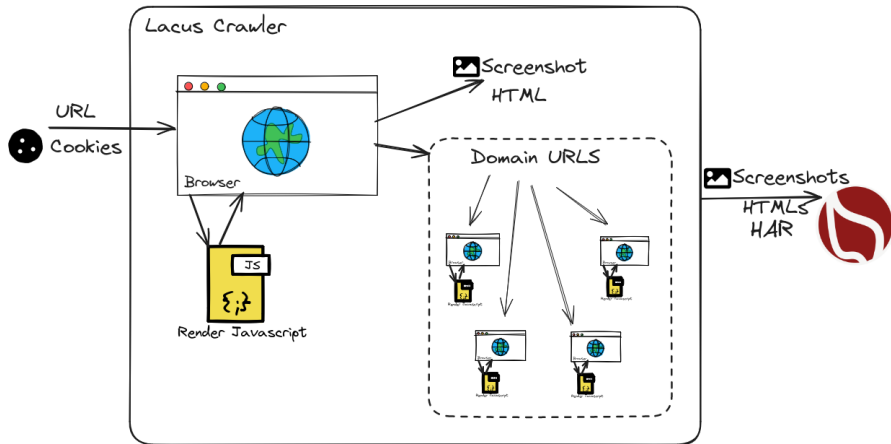
Crawler: Cookiejar

Use your cookies to login and bypass captcha

🐛 Edit Cookiejar 🗑

Description	Date	UUID	User
3thxemke2x7hcibu.onion	2020/03/31	90674deb-38fb-4eba-a661-18899ccb3841	admin@admin.test

Edit Description ✏️   **Add Cookies** 🍪

🗑 ✏️
```
{
    "domain": ".3thxemke2x7hcibu.onion",
    "name": "mybb[lastactive]",
    "path": "/forum/",
    "value": "1583829465"
}
```

🗑 ✏️
```
{
    "domain": ".3thxemke2x7hcibu.onion",
    "name": "loginattempts",
    "path": "/forum/",
    "value": "1"
}
```

🗑 ✏️
```
{
    "domain": ".3thxemke2x7hcibu.onion",
    "name": "sid",
    "path": "/forum/",
    "value": "047ab0cd97ff5bcc77edb6a"
}
```

🗑 ✏️
```
{
    "name": "remember_token",
    "value": "12|58cddd1511d74d341f23"
}
```

🗑 ✏️
```
{
    "domain": ".3thxemke2x7hcibu.onion",
    "name": "mybb[announcements]",
    "path": "/forum/",
    "value": "0"
}
```

Crawler: Cookiejar

Collection - automate collection

- Collecting data from various chat sources can be a **tedious task for analysts**.
- AIL offers a set of feeders (e.g., Telegram, Discord, etc.) that can be used to subscribe to chat channels.
- All the **collected messages are then processed and analyzed** within the AIL's *processing* and *analysis* stages.

Processing - OCR

Иринка
Всё,что летит из рашки
ПУСКАЙ ЛЕТИТ ОБРАТНОЩ
УМНОЖИТСЯ СТОКРАТНОЮ!
ПУСТЬ ВАШЕ БУДЕТ ВАМЩ!
ДА БУДЕТ ТАКІ
01.01

Irinka
Everything that flies from Russia
LET IT FLY BACK
IT WILL MULTIPLY A HUNDRED TIMES!
LET YOURS BE YOURS!
LET IT BE SO
01.01

🆎 ru

Tags:

AIL Framework: Features

- Extracting **credit card numbers, credentials, phone numbers, ...**
- Extracting and validating potential **hostnames**
- Keeps track of **duplicates**
- Submission to threat sharing and incident response platforms (**MISP** and **TheHive**)
- **Full-text indexer** to index unstructured information
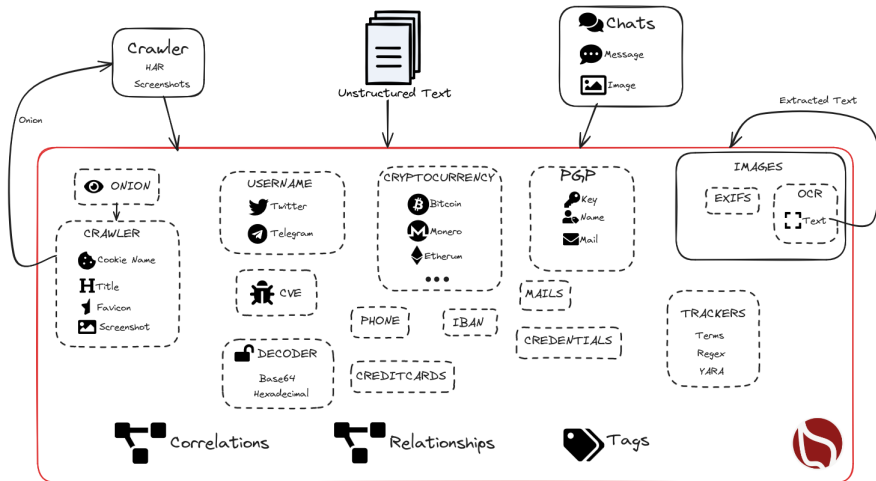- **Tagging** for classification and searches
- Terms, sets, regex, and YARA **tracking, occurrences, and history**
- Archives, files, and raw **submission** from the UI
- Correlation engine based on PGP ID, cryptocurrencies, decoded (Base64, ...), usernames, cookie names, and many selectors to find relationships
- And many more

Trackers - Retro Hunt

- Search and monitor specific keywords/patterns
  - Automatic tagging
  - Email/webhook notifications
- Track Word
  - ddos
- Track Set
  - booter, ddos, stresser; 2
- Track Regex
  - circl\lu
- YARA rules
  - https://github.com/ail-project/ail-yara-rules

Live Demo

YARA Tracker

Retro Hunt

test

✅ completed

🗑️

Date	2023/05/10
Description	None
Tags	
Creator	admin@admin.test
Filters	{ "item": { "date_from": "20230304", "date_to": "20230601" } }
Objects Match	item 3

🔍 Show Objects

```
rule certificates
{
    meta:
        author = "@KevTheHermit"
        info = "Part of PasteHunter"
        reference = "https://github.com/kevthehermit/PasteHunter"

    strings:
        $ssh_priv = "BEGIN RSA PRIVATE KEY" wide ascii nocase
        $openssh_priv = "BEGIN OPENSSH PRIVATE KEY" wide ascii nocase
        $dsa_priv = "BEGIN DSA PRIVATE KEY" wide ascii nocase
        $ec_priv = "BEGIN EC PRIVATE KEY" wide ascii nocase
        $pgp_priv = "BEGIN PGP PRIVATE KEY" wide ascii nocase
        $pem_cert = "BEGIN CERTIFICATE" wide ascii nocase
        $pkcs7 = "BEGIN PKCS7"

    condition:
        any of them
}
```
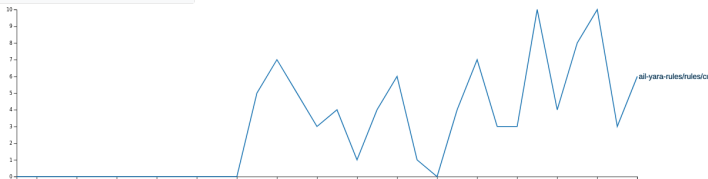
Show 10 ⇕ entries   Search:

Type ⇅			Id	Tags		
🟣			archive/gist.github.com/2023/04 /14/luizmiranda7_3b3d1133a3d3842092c5fc5fb39e84f2.gz	infoleak:automatic-detection="private-key" test23 test12 infoleak:automatic-detection="certificate"		
🟣			submitted/2023/04/20/submitted_cc9190ab-80d2-4d2b-9c9e-97c51e69a855.gz	infoleak:submission="manual" test12 infoleak:automatic-detection="rsa-private-key" infoleak:automatic-detection="vpn-static-key" test23 infoleak:automatic-detection="certificate" infoleak:automatic-detection="onion"		
🟣			archive/gist.github.com/2023/04 /13/chipzoller_d8d6d2d737d02ad4fe9d30a897170761.gz	test12 test23 infoleak:automatic-detection="certificate"		

Showing ... 3 entries

Correlations and relationship
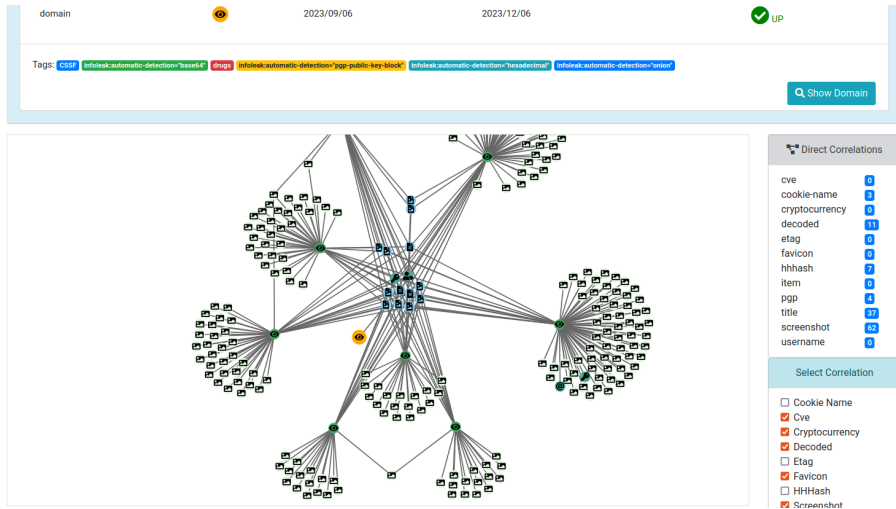
Investigations

Tor Coin Mixer

UUID	9189d0e7c04c47a29f85666e9507e0a5
Creator	admin@admin.test
Tags	dark-web:topic="mixer"
Date	2023-05-31
Threat Level	medium
Analysis	initial
Info	Tor Coin Mixer
# Objects	6
Timestamp	2023-05-31 12:50:45
Last change	2023-05-31 12:54:20

Delete **Edit** **Export as Event**

Objects

Show 10 entries

Search:

Type	Id	Tags	
onion	jambleny7zgpxknhjtmj3rnhfdajmyddqxbuhf6voa32h5w4otux3crqd.onion	infoleak:automatic-detection="onion" infoleak:automatic-detection="pgp-public-key-block"	
onion	bitmixhft4cpncluhwflussk23itvowswbe4tlrdree74oxjmz2vyqqd.onion	infoleak:automatic-detection="onion"	
key	0xD3B280956F0E7CAF		
mail	support@jambler.io		
telegram	jambler		
name	Jambler.io		

Showing 1 to 6 of 6 entries

Previous **1** Next

AIL Framework: Extensible Capabilities

- Extending AIL to add a new **analysis module** can be done in 50 lines of Python.
- The framework **supports multi-processors/cores by default**. Any analysis module can be started multiple times to support faster processing during peak times or bulk import.
- **Multiple** concurrent **data inputs**.
- Tor Crawler (handles cookies and authentication).
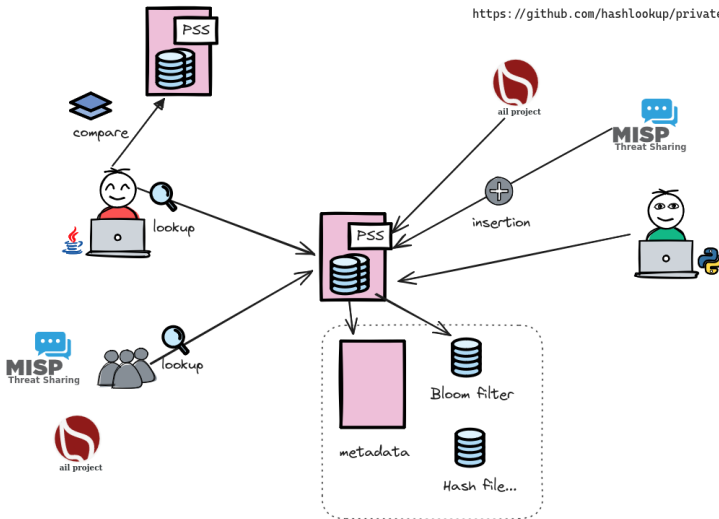- Feeders: Discord, Telegram, ...

Ongoing Developments

- Two-Factor Authentication (2FA) - One-Time Password (OTP)
- Bloom Filter Filtering - Private Search Set (PSS)
- Relationships and activity between chats: message forwards, replies, etc.

Private Search Set (PSS)

Private Search Set (PSS)
https://github.com/hashlookup/private-search-set/

Links

- AIL project website https://ail-project.org
- AIL project open source framework
  https://github.com/ail-project
- Training materials
  https://github.com/ail-project/ail-training
- Online chat https://gitter.im/ail-project/community

ail project

MISP - LEA

- Law Enforcement Agency Information Sharing Community
- AIL-LEA instance hosted by CIRCL
- Request access by sending an email to **info@misp-lea.org** or **info@circl.lu**
- Request free training at **info@misp-lea.org** or **info@circl.lu**