



**CIRCL**  
Computer Incident  
Response Center  
Luxembourg



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

CIRCL - Virtual Summer School 2025

🏠 <https://ail-project.org/>

---

Alexandre Dulaunoy - [alexandre.dulaunoy@circl.lu](mailto:alexandre.dulaunoy@circl.lu)

Aurelien Thirion - [aurelien.thirion@circl.lu](mailto:aurelien.thirion@circl.lu)

July 17, 2025

CIRCL <https://www.circl.lu>

## Background

- Over the past years, CIRCL has developed the AIL project<sup>1</sup> 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<sup>2</sup> comprising various modules designed for the **collection, crawling, digging, and analysis of unstructured data**.

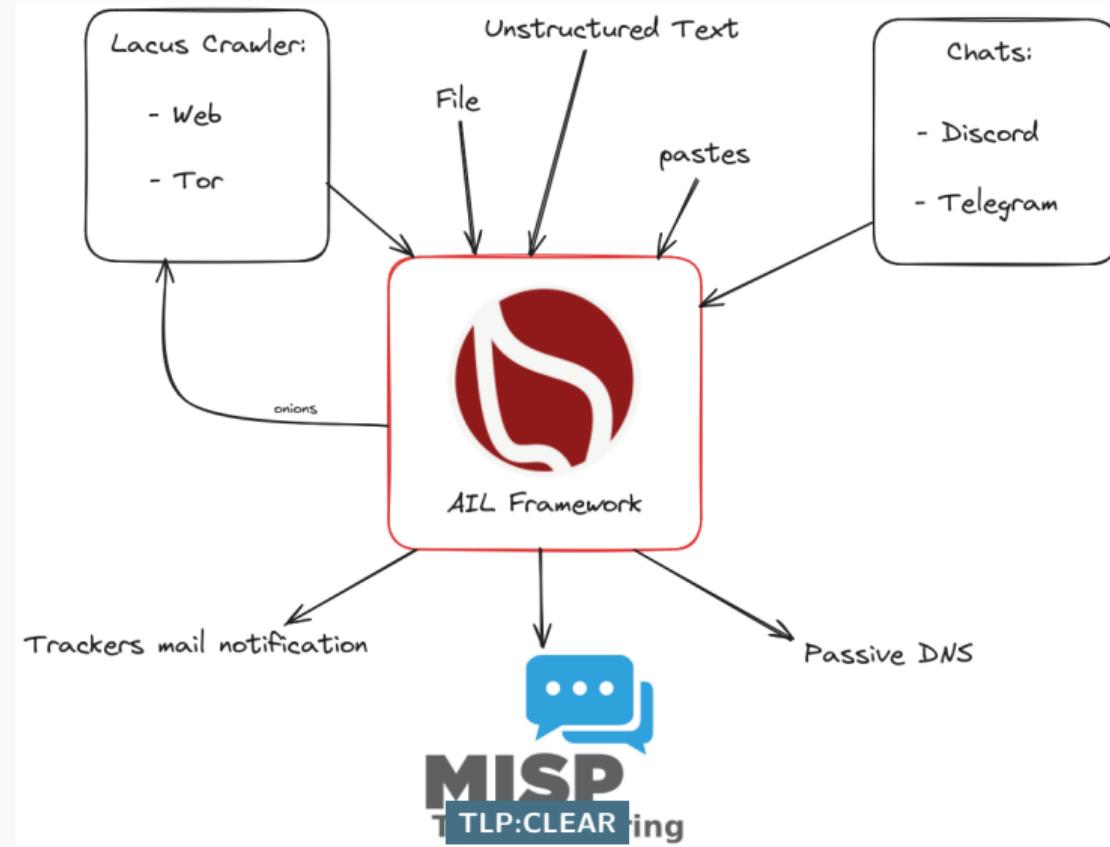


---

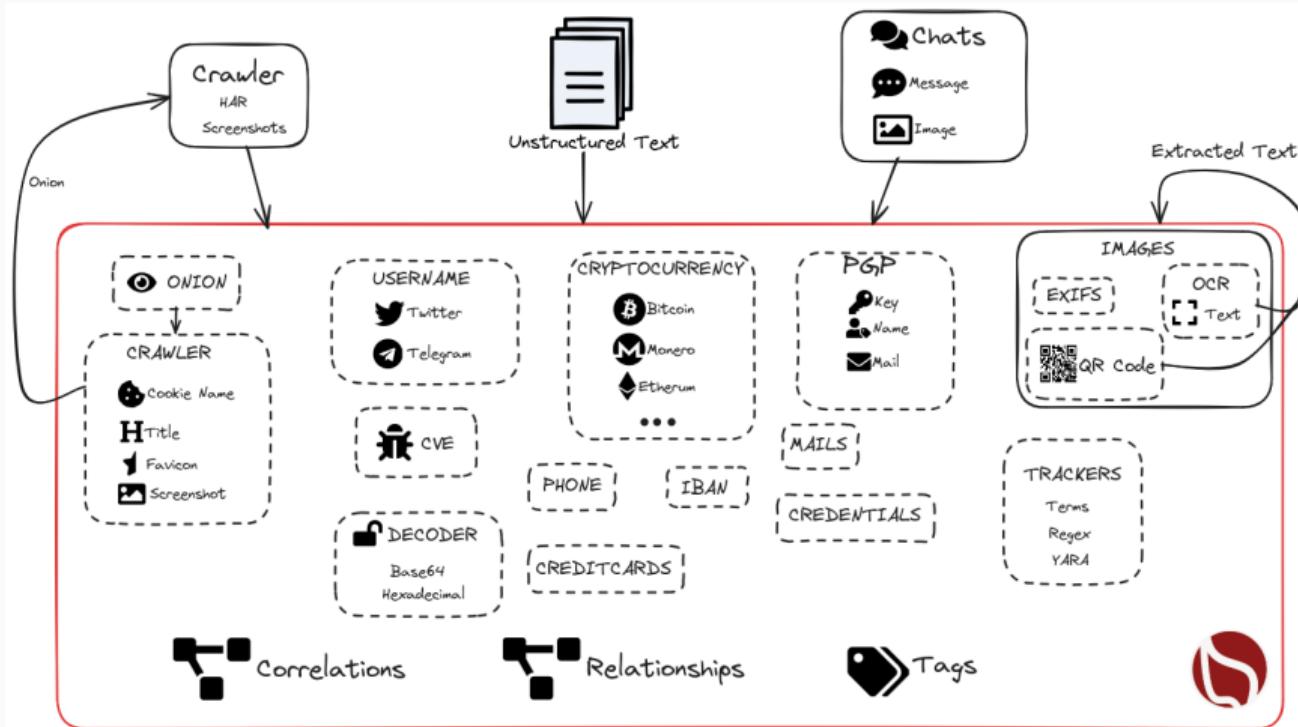
<sup>1</sup><https://www.ail-project.org/>

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

# High Level Overview

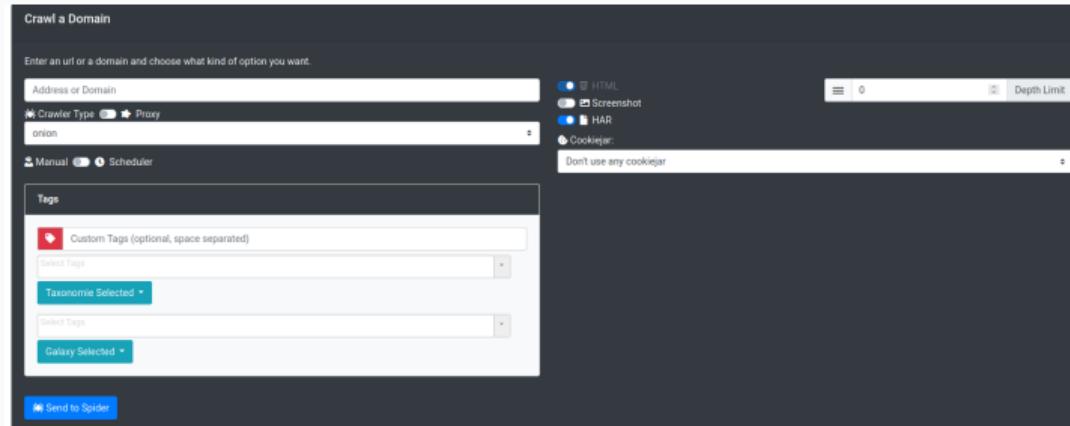


# Analysis of unstructured information



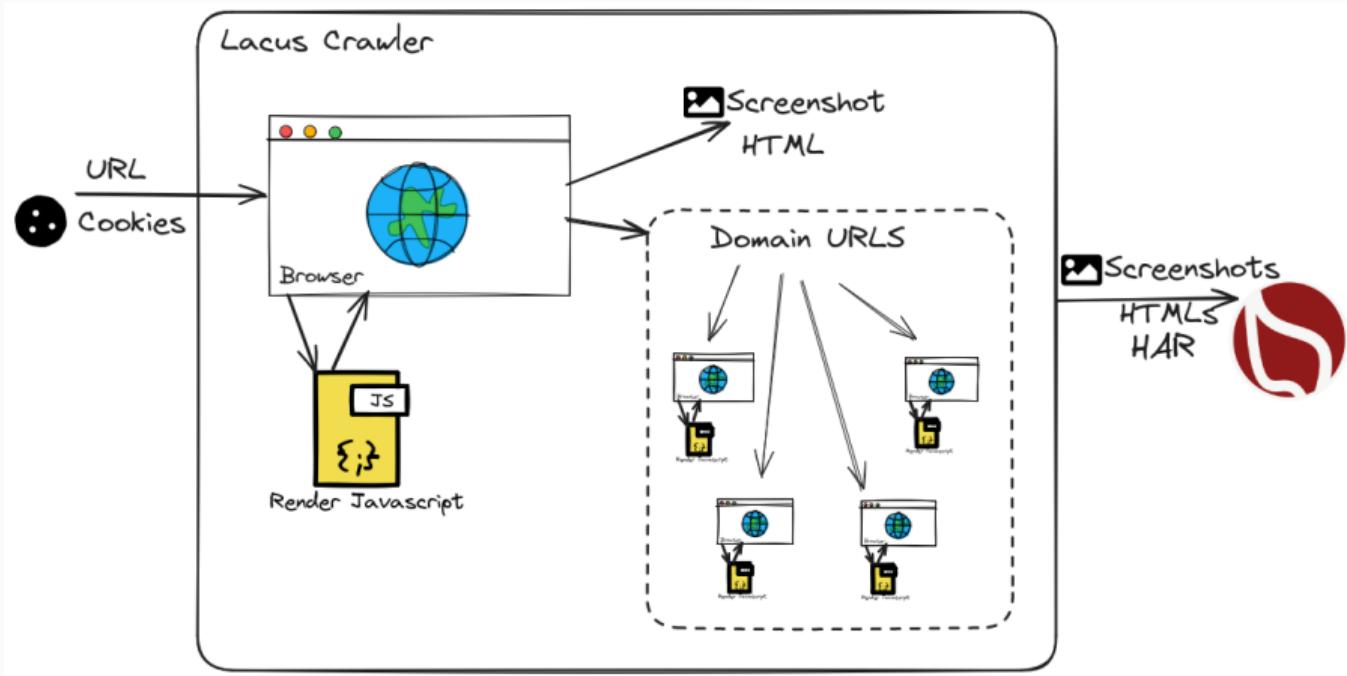
## Collection - automate crawling

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



<sup>3</sup><https://www.ransomlook.io/>

# Collection - Lacus Crawler<sup>4</sup>



<sup>4</sup><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](#)

[Edit](#) [Delete](#)

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

[Edit](#) [Delete](#)

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

[Edit](#) [Delete](#)

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

[Edit](#) [Delete](#)

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

[Edit](#) [Delete](#)

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

TLP:CLEAR

7/26

# Crawler: Cookiejar

3thxemke2x7hcibu.onion :

DOWN

First Seen	Last Check	Ports
2020/03/09	2020/03/30	[80]

infoleak:automatic-detection="onion" infoleak:automatic-detection="base64"

manual

Show Domain Correlations 139

Add to MISP Export

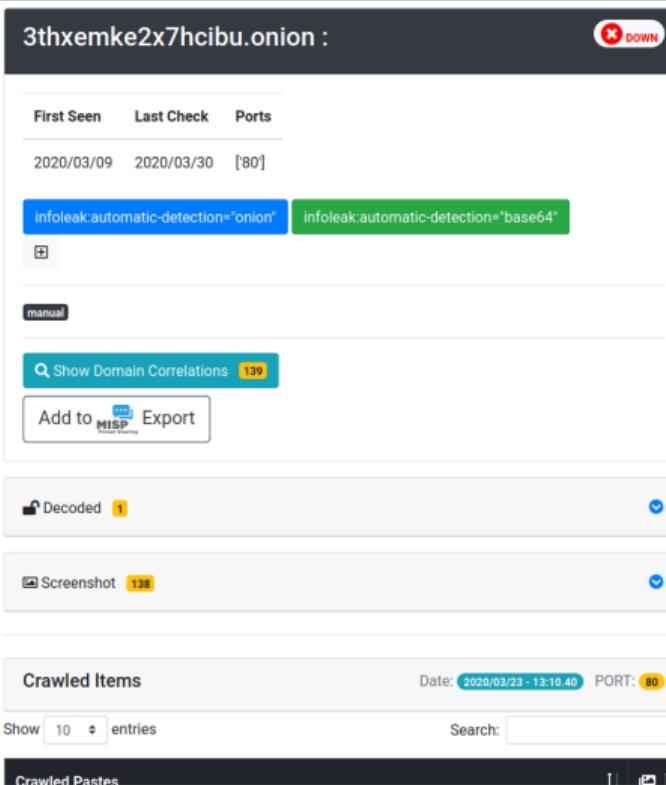
Decoded 1

Screenshot 138

Crawled Items Date: 2020/03/23 - 13:10:40 PORT: 80

Show 10 entries Search:

Crawled Pastes



Hide Full resolution

## Shere Khan

Welcome back, zuipori. You last visited: 03-20-2020, 01:35 PM Log Out

User CP View New Posts View Today's Posts Private Messages (Unread 2, Total 2)

You have 2 unread private messages. The most recent is from Jok3 titled KEY FOR PRIVATE SECTIONS

Shere Khan - Official Forum  
Private Messages

Menu User CP Home Messenger Compose Unread Sent Items Drafts Trash Can Tracking Self Folders Your Profile Edit Profile Change Password Change Email Change Avatar Change Signature Edit Options Miscellaneous Group Memberships Add/Ignore List Manage Attachments Send Drafts Subscribed Threads Forum Subscriptions View Profile

Inbox Enter Keywords Search PMs (Advanced Search) Message Title Sender Date/Time Sent [asc]

Message Title	Sender	Date/Time Sent [asc]
KEY FOR PRIVATE SECTIONS	Jok3	3 hours ago
Verification	Jok3	03-09-2020, 11:55 AM

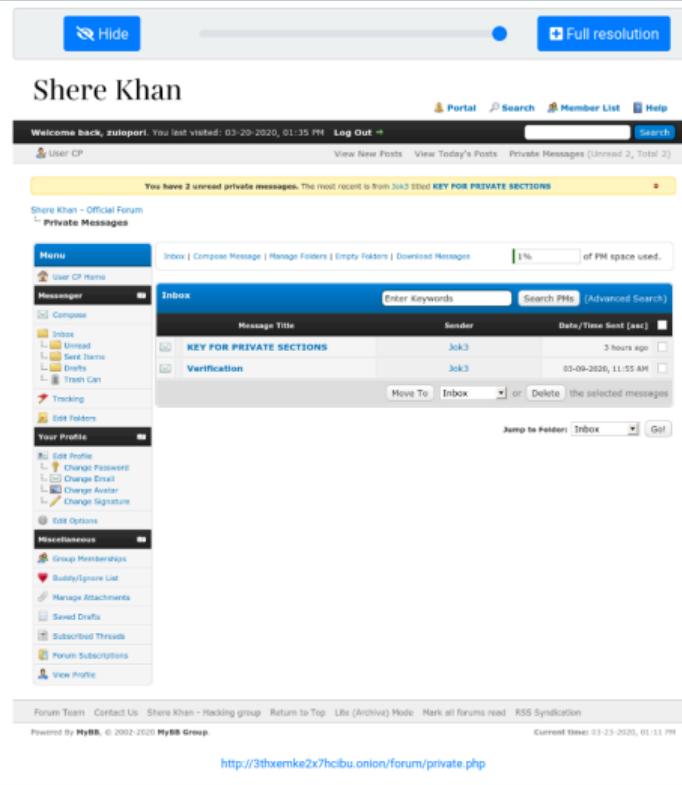
Move To Inbox or Delete the selected messages

Jump to Folders: Inbox Get

Forum Team Contact Us Shere Khan - Hacking group Return to Top Lite (Archive) Mode Mark all forums read RSS Syndication

Powered By MyBB, © 2002-2020 MyBB Group Current Time: 03-23-2020, 01:11 PM

<http://3thxemke2x7hcibu.onion/forum/private.php>



TLP:CLEAR

8/26

# 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) that can be used to subscribe to chat channels.
- All the **collected messages are then processed and analyzed** within AIL's *processing* and *analysis* stages.

DDoSia Project :



Name	ID	Created at	First Seen	Last Seen	NB Sub-Channels	Participants
DDoSia Project	2125229770	2024-03-07 09:03:02	2024-03-07	2024-04-12	6	695

Tags:

[Investigations](#) [Correlations](#)

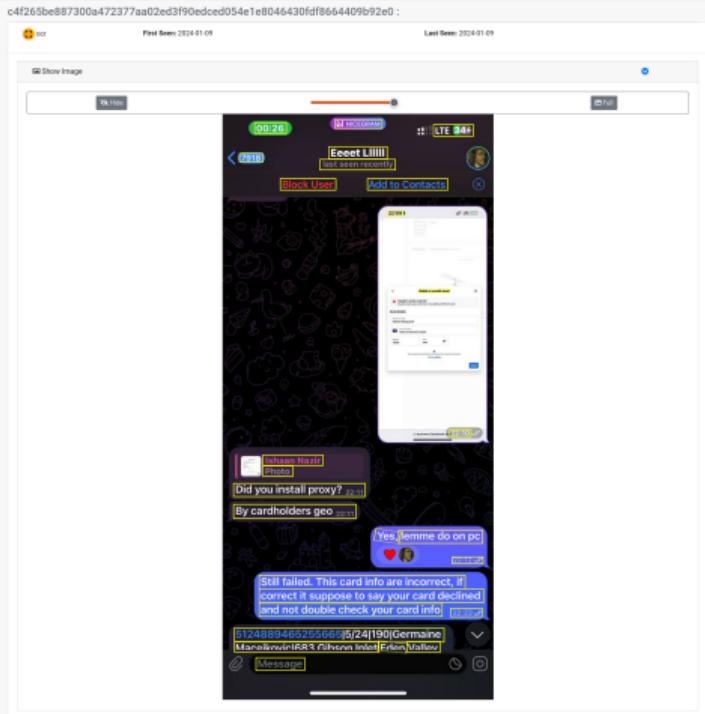
## Sub-Channels:

Show [10](#) entries

Search:

Icon	Name	ID	Created at	First Seen	Last Seen	
	General	2125229770/1	2024-03-07 06:43:47	2024-03-07	2024-04-12	4121
	Поддержка - отвечаем на вопросы Support - answering questions	TLP:CLEAR	2024-03-07 10:00:24	2024-03-07	2024-04-12	2264
	English support	2125229770/138	2024-03-07 08:03:02	2024-03-07	2024-04-11	297

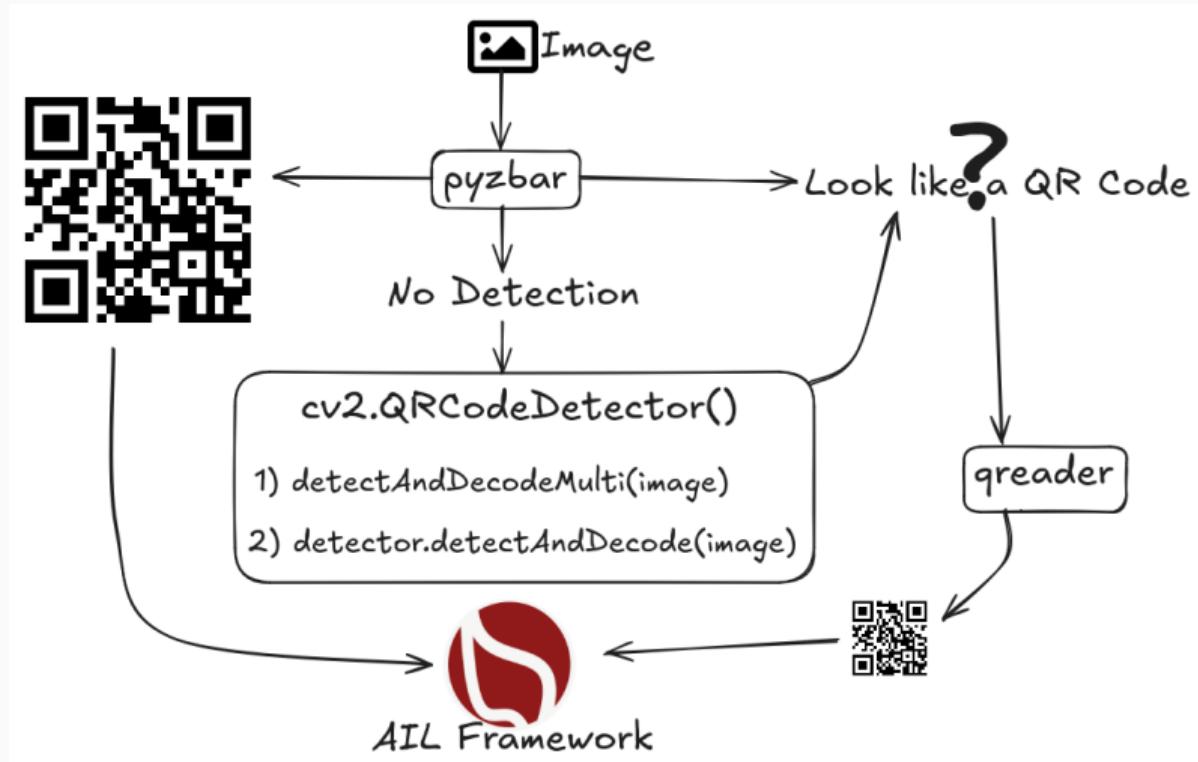
# OCR: Optical Character Recognition



- Threat actors are often verbose and frequently share extensive details in private channels.
- Many messages contain screenshots and images.
- Text detection and extraction are performed across 80+ languages using a CRNN (Convolutional Recurrent Neural Network).
- Enables keyword-based matching and detection.

TLP:CLEAR

# QR Code Extractor



# QR Codes

blue\_forever20  
2022-12-16  
16:51:04



likeaboss

000201010211021643327300096494304155290090009649415312676076426760764000  
Mai62360520AP9J3I00000000000000000708447583166304F69D

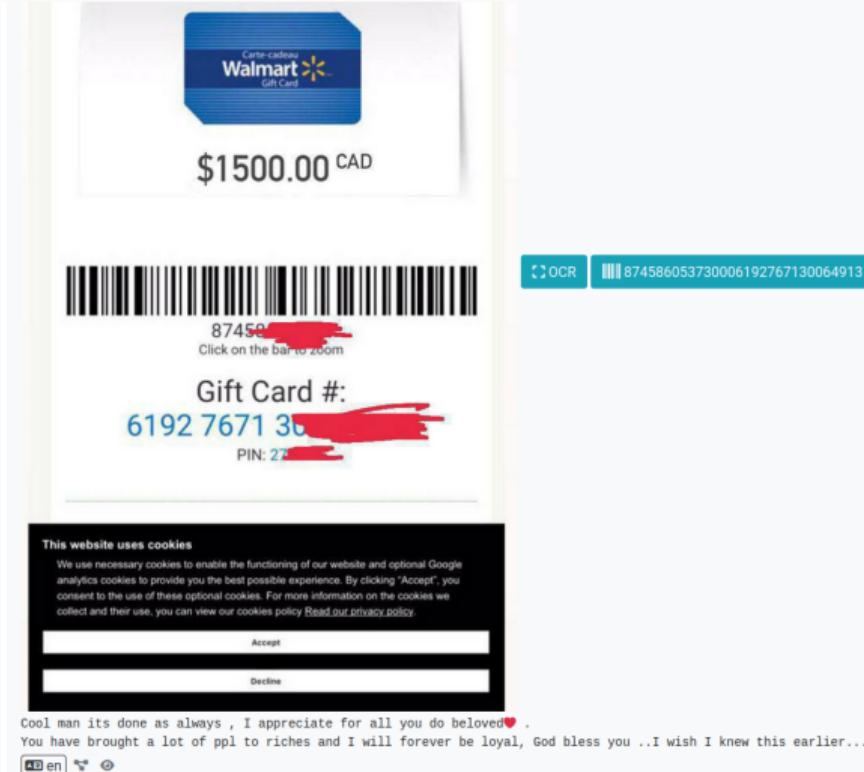
Someone with union pay link or QR code  
Hit me up

We can make money

**TLP:CLEAR**

en

# Bar Codes



TLP:CLEAR

- Extracting **credit card numbers, credentials, phone numbers, ...**
- Extracting and validating potential **hostnames**
- Submission to threat sharing and incident response platforms (**MISP** and **FlowIntel<sup>5</sup>**)
- **Tagging<sup>6</sup>** 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

---

<sup>5</sup><https://github.com/flowintel/flowintel>

<sup>6</sup>Relying on MISP taxonomies and galaxy

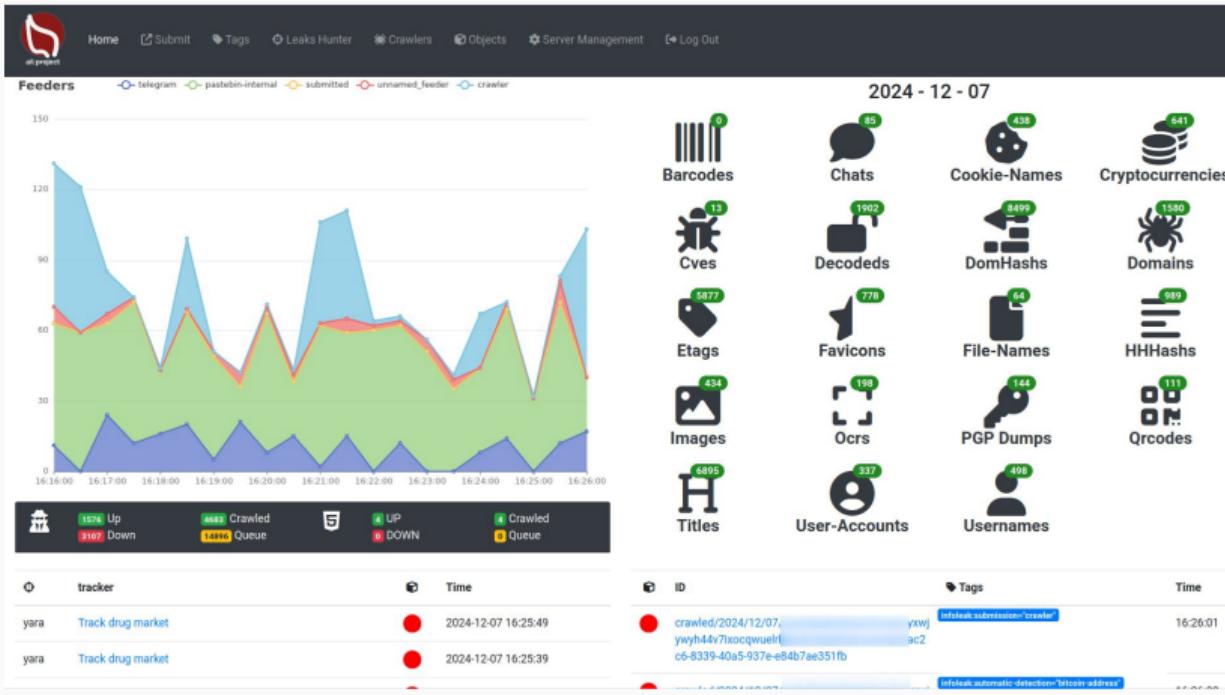
# Live Trackers & Retro Hunt

- 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

---

# Dashboard



TLP:CLEAR

16/26

# Search

The screenshot shows a search interface with several sections:

- Tor and Web Search:** A search bar with "Tor" selected. Options include "Type: Tor", "Web", and "All". A search input field contains "content to Search" with a search button and an information icon.
- Search Domain by name:** A search bar for "Domain name" with a "Domain" toggle switch set to "Orion Domains". There is also a "Web Domains" toggle switch.
- Titles Search:** A search bar for "Content Search" with a "Case Sensitive" toggle switch.
- Chats Search:** A search bar for "discord" with "Type: discord", "matrix", "telegram", and "All Chats" options. A search input field contains "content to Search" with a search button and an information icon.
- Usernames Search:** A search bar for "Usernames" with a placeholder "Usernames to Search".

TLP:CLEAR

17/26

# YARA Tracker

Certificate

Type: **yara**

Tracked: [all-yara-rules/rules/crypto/certificate.yar](#)

Date: 2023/05/12

Level: Global

Creator: admin@admin.test

First Seen: 2023 / 05 / 12

Last Seen: 2023 / 05 / 31

Tags:

Mails:

Webhook:

Filters: [No filters](#)

Objects Match: **decoded 6**  
[Item 88](#)

[Edit Tracker](#)

**Yara Rule:**

```
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
        $sec_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
}
```

2023-05-12 | 2023-05-31

[Tracked Objects](#)

TLP:CLEAR

all-yara-rules/rules/c

# 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
}
```

Type	Id	Tags
●	archive/gist.github.com/2023/04/14/luzmiranda_3b3d1133a3d3842092c5fc5fb39e84f2.gz	infoleak:automatic-detection="private-key" test12 test12 infoleak:automatic-detection="certificate"
●	submitted/2023/04/20/submitted_cc9190ab-80d2-4d2b-9c0e-97c51e69a855.gz	infoleak:submission="manual" test12 infoleak:automatic-detection="rsa-private-key" infoleak:automatic-detection="vpn-static-key" test12 infoleak:automatic-detection="certificate" infoleak:automatic-detection="onion"
●	archive/gist.github.com/2023/04/13/chipzoller_d8d6d2d737d02ad4fe9d30a897170761.gz	test12 test12 infoleak:automatic-detection="certificate"

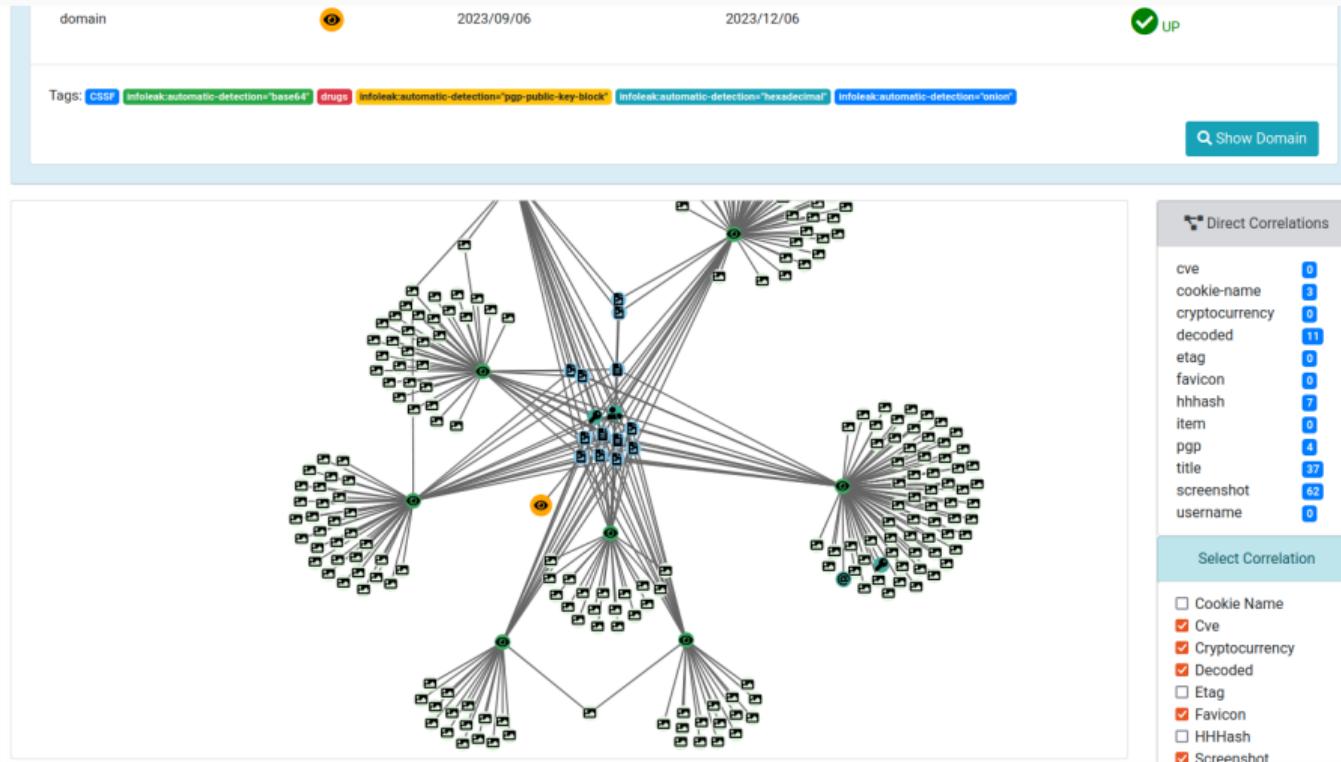
Showing 1 to 3 of 3 entries

Previous 1 Next

TLP:CLEAR

19/26

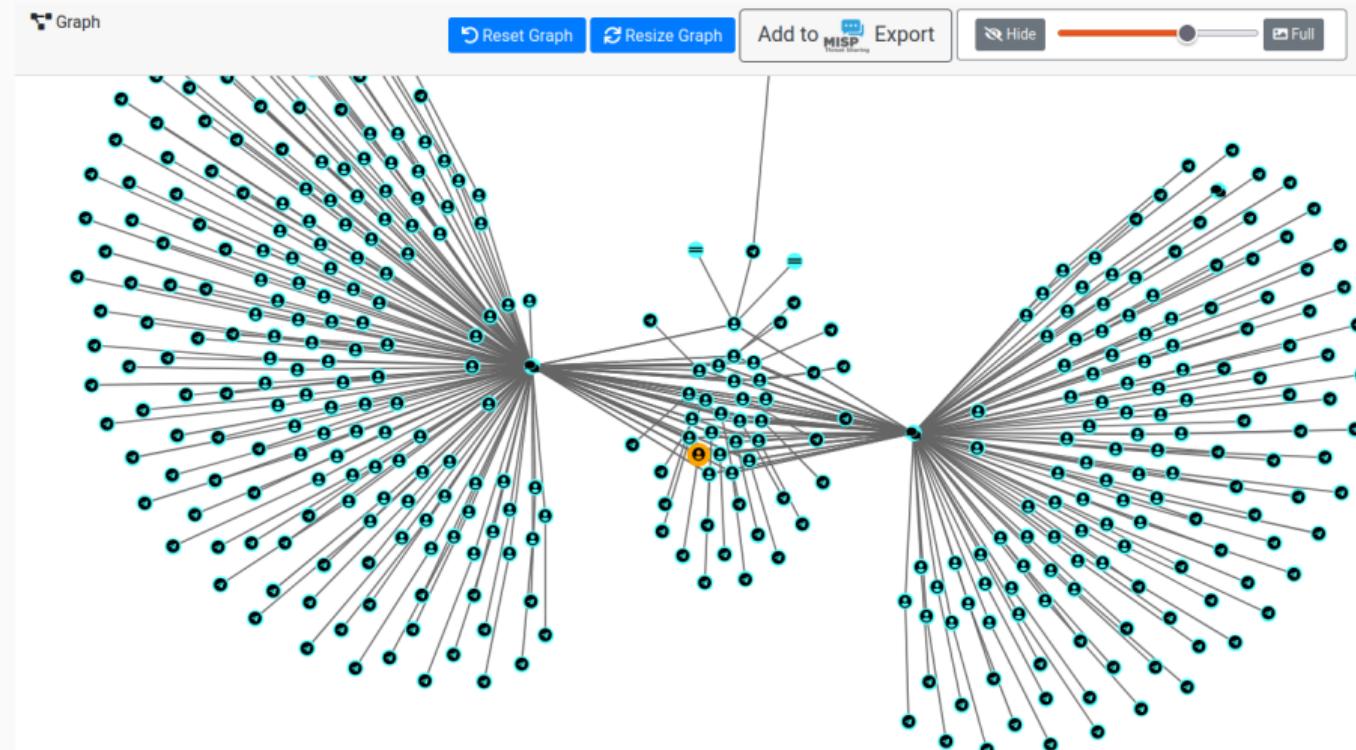
# Correlations and relationship



TLP:CLEAR

20/26

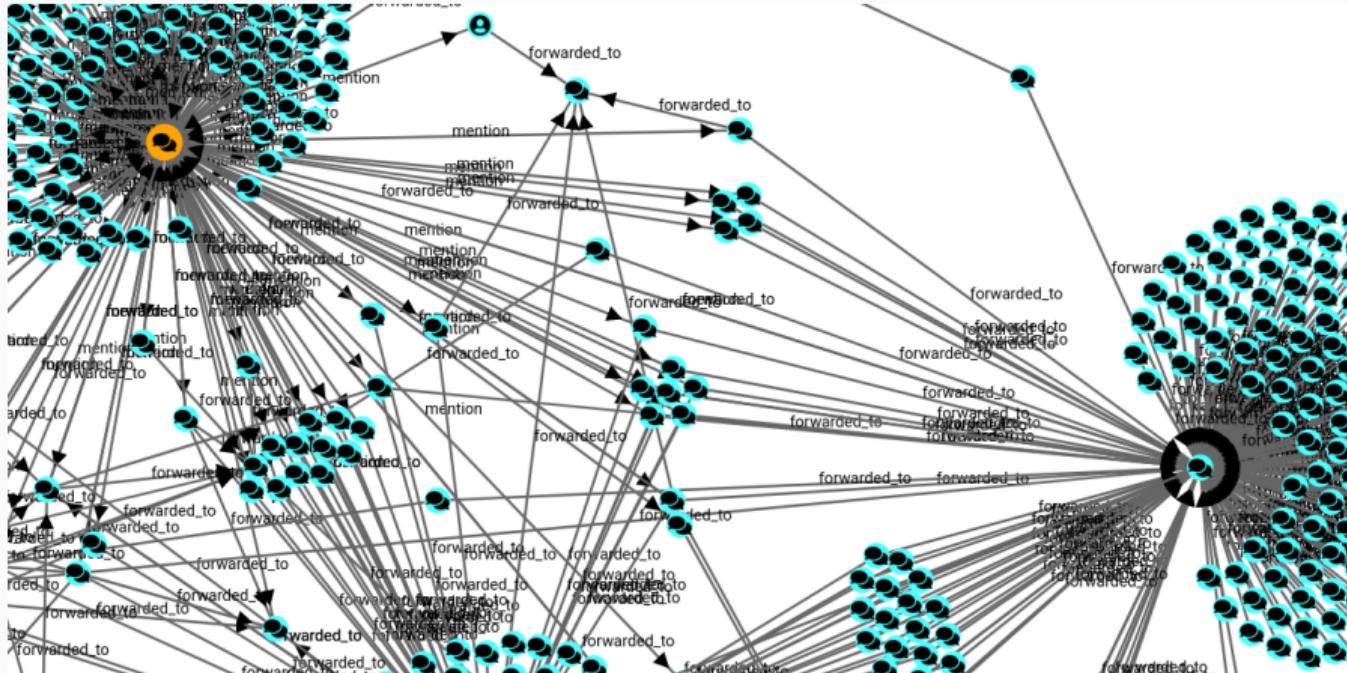
# User Correlation - Common Chats



TLP:CLEAR

21/26

# What are the Relationships Between Chats Group?



# Investigations

### Tor Coin Mixer

UUID	9189d0e7c04c47a29f85666e9507e0a5	<a href="#">Delete</a>	<a href="#">Edit</a>	<a href="#">Export as Event</a>
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			

### Objects

Show 10 entries Search:

Type	Id	Tags	
onion	jambler1y2zp8knhjbnj3mhfdajmyddqbxuf1voa32h5w4o6ux3cqrd.onion	[info:automatic-detections="index"] [info:automatic-detections="pgp-public-key-block"]	<a href="#">Delete</a>
onion	bitmonhf4cpncluhwfifuskk23tvowswe4tlthree74oxjmz2yyqqd.onion	[info:automatic-detections="index"]	<a href="#">Delete</a>
key	0x0B280956F0E7CAF		<a href="#">Delete</a>
mail	support@jambler.io		<a href="#">Delete</a>
telegram	jambler		<a href="#">Delete</a>
name	Jambler.io		<a href="#">Delete</a>

Showing 1 to 6 of 6 entries

TLP:CLEAR

Previous [1](#) Next

23/26

# 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

- **Mail Search – Next Release**
- **Lacus crawler improvements:** proxy selection, local cache, and cookie state transfer
- **Translate search:** support for searching in other languages
- **Advanced video processing and extraction**
- **MISP export with new correlation types**
- **Automatic geolocation**

## 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>



## Thank you for your attention

- AIL project<sup>7</sup> : <https://github.com/ail-project/ail-framework>
- For questions, contact: [info@circl.lu](mailto:info@circl.lu)

---

<sup>7</sup>All techniques and indicators mentioned in these slides are implemented in the AIL project, using an instance backed by a three-year dataset collected from Tor hidden services and various social networks.