CIRCL
Computer Incident
Response Center
Luxembourg

ail project

# 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
Aurelien Thirion - aurelien.thirion@circl.lu

July 17, 2025

CIRCL https://www.circl.lu

TLP:CLEAR

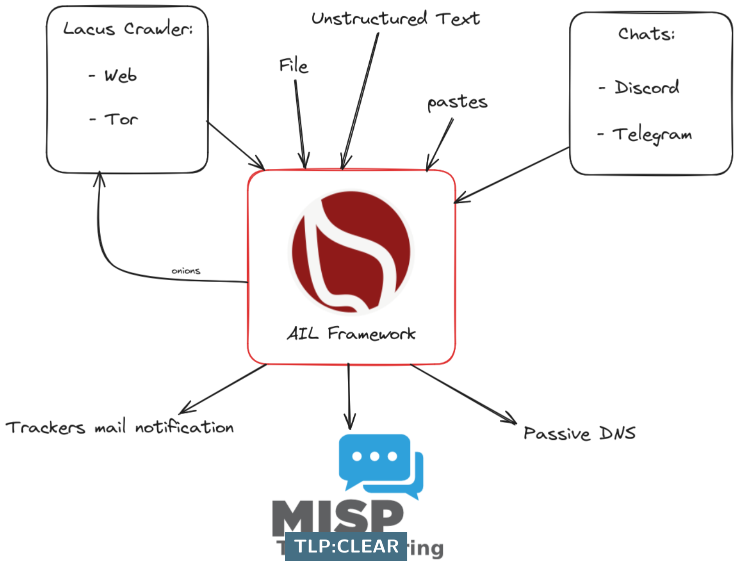
# Background

- 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

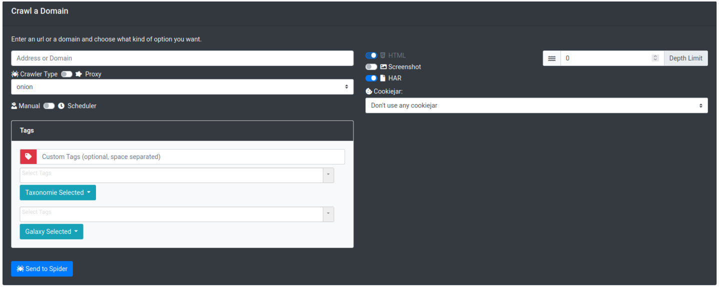
# High Level Overview



TLP:CLEAR

# *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**.



Crawl a Domain

Enter an url or a domain and choose what kind of option you want.

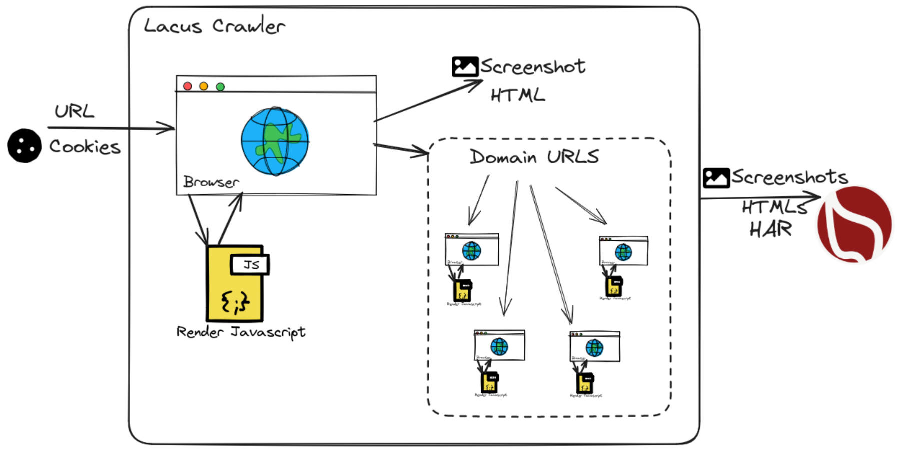
onion

Cookiejar: Don't use any cookiejar

Depth Limit: 0

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

# *Collection* - **Lacus Crawler**[4]

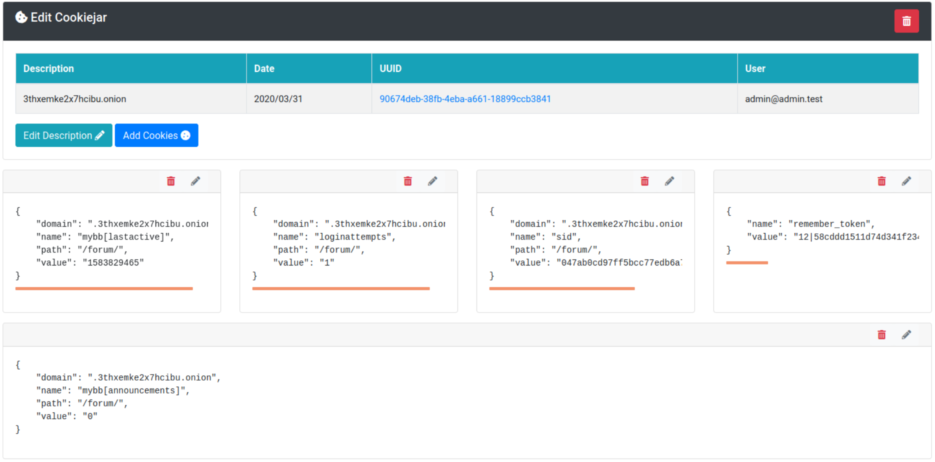


[4] https://github.com/ail-project/lacus

TLP:CLEAR

# Crawler: Cookiejar

Use your cookies to login and bypass captcha



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

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

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

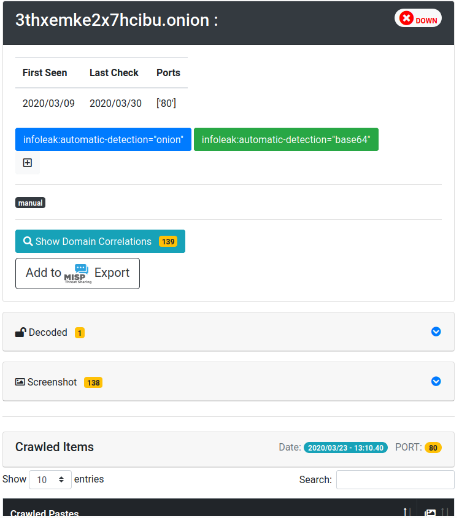
```
{
    "domain": ".3thxemke2x7hcibu.onio
    "name": "sid",
    "path": "/forum/",
    "value": "047ab0cd97ff5bcc77edb6a
}
```

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

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

TLP:CLEAR

# Crawler: Cookiejar



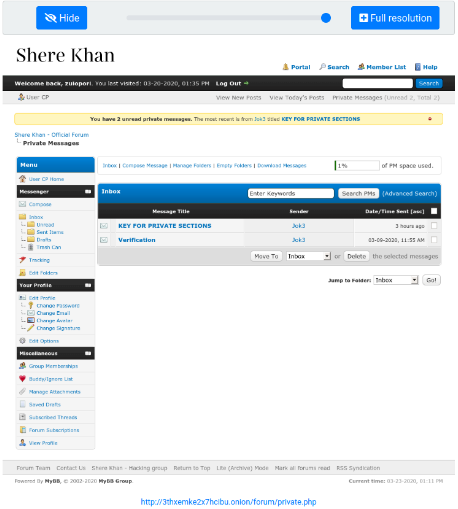
## 3thxemke2x7hcibu.onion :

| 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 Threat Sharing Export
Decoded 1
Screenshot 138

Crawled Items — Date: 2020/03/23 - 13:10.40 PORT: 80



Shere Khan

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

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

1% of PM space used.

Inbox

| 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 Folder: Inbox Go!
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.

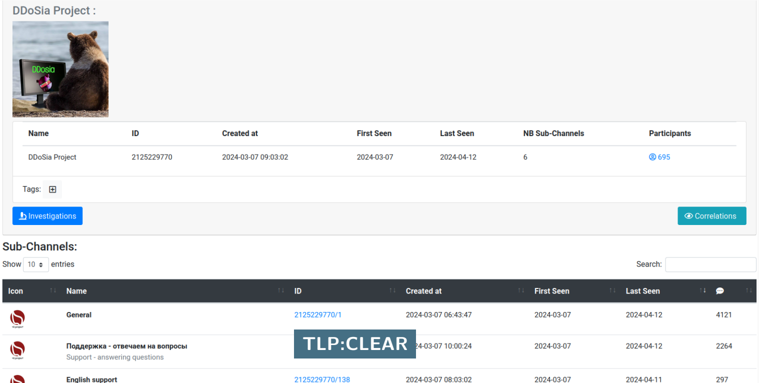
Current time: 03-23-2020, 01:11 PM

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

TLP:CLEAR

# 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: ⊞

Sub-Channels:

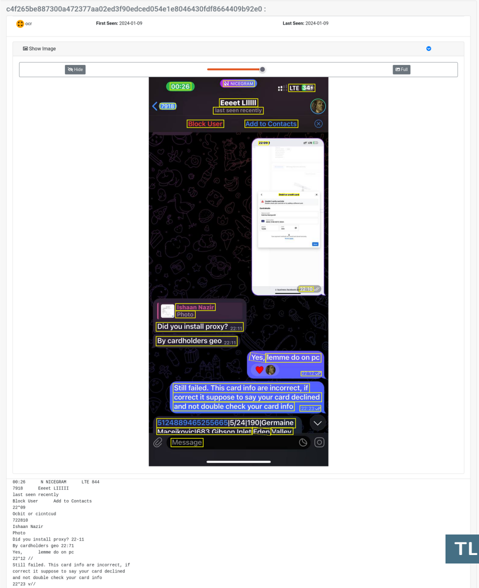
Show 10 entries

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<br>Support - answering questions | | 24-03-07 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 |

TLP:CLEAR

# OCR: Optical Character Recognition



c4f265be887300a472377aa02ed3f90edced054e1e8046430fdf8664409b92e0 :

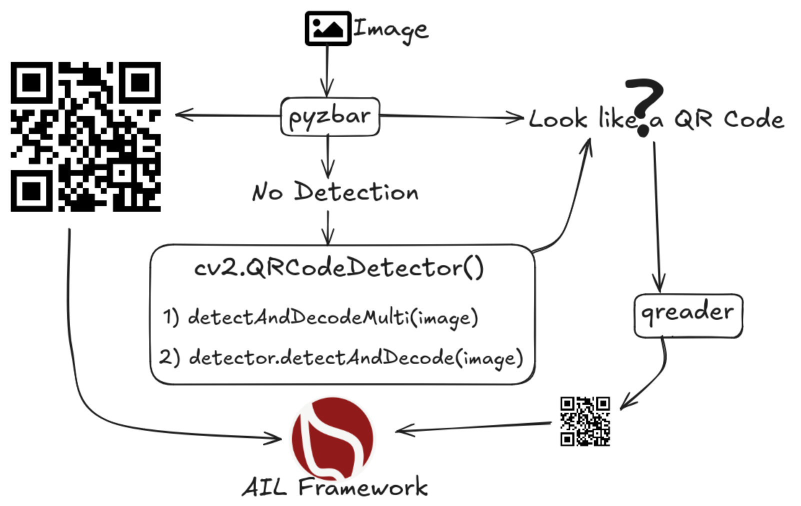
ocr **First Seen:** 2024-01-09 **Last Seen:** 2024-01-09

```
00:26      N NICEGRAM      LTE 844
7918      Eeeet LIIIII
last seen recently
Block User      Add to Contacts
22*09
Ocbit or cicntcud
722810
Ishaan Nazir
Photo
Did you install proxy? 22-11
By cardholders geo 22:71
Yes,      lemme do on pc
22*12 //
Still failed. This card info are incorrect, if
correct it suppose to say your card declined
and not double check your card info
22*23 v//
```

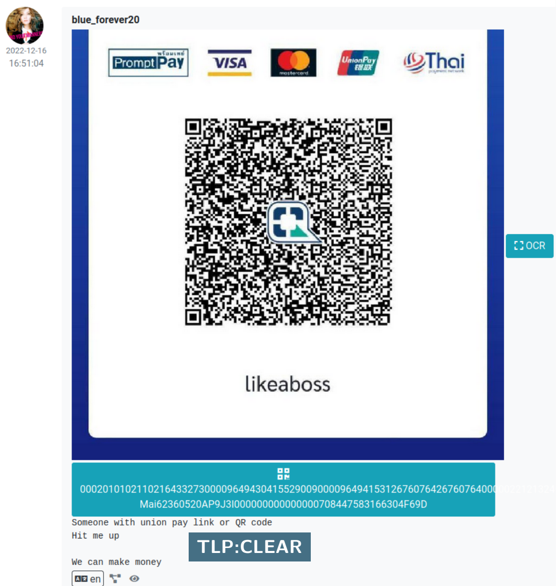
- Threat actors are often verbose and frequently **share extensive details in private channels**.
- Many messages contain screenshots and images.
- Text detection and extraction are performed across 80+ languages using a CRNN (Convolutional Recurrent Neural Network).
- **Enables keyword-based matching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

TLP:CLEAR

# QR Code Extractor



# QR Codes



blue_forever20

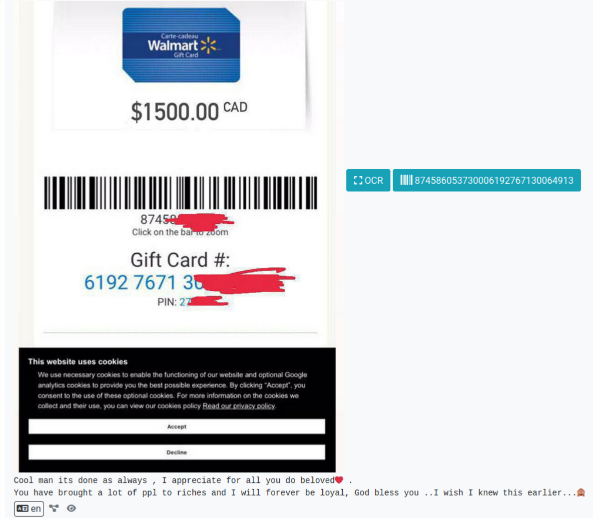
2022-12-16
16:51:04

0002010102110216433273000096494304155290090000096494153126760764267607640000022121324
Mai62360520AP9J3I0000000000000000708447583166304F69D

```
Someone with union pay link or QR code
Hit me up

We can make money
```

TLP:CLEAR

# Bar Codes



$1500.00 CAD

8745[redacted]
Click on the bar to zoom

Gift Card #:
6192 7671 3[redacted]
PIN: 27[redacted]

**This website uses cookies**

We use necessary cookies to enable the functioning of our website and optional Google analytics cookies to provide you the best possible experience. By clicking "Accept", you consent to the use of these optional cookies. For more information on the cookies we collect and their use, you can view our cookies policy Read our privacy policy.

87458605373000619276713006491 3

```
Cool man its done as always , I appreciate for all you do beloved❤️ .
You have brought a lot of ppl to riches and I will forever be loyal, God bless you ..I wish I knew this earlier...🙈
```

TLP:CLEAR

# AIL Framework: Features

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

[5] https://github.com/flowintel/flowintel

[6] Relying on MISP taxonomies and galaxy

TLP:CLEAR

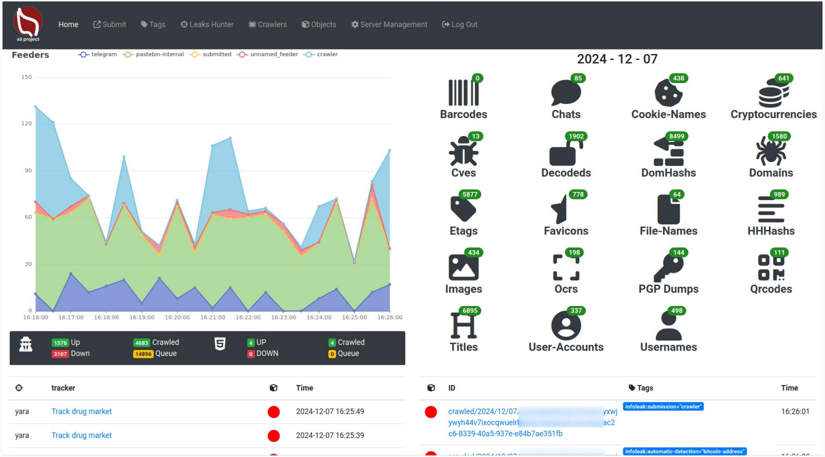
# Live Trackers & Retro Hunt

- 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`

TLP:CLEAR

# Live Demo

# Dashboard



2024 - 12 - 07

| | | | |
|---|---|---|---|
| 0 Barcodes | 85 Chats | 438 Cookie-Names | 641 Cryptocurrencies |
| 13 Cves | 1902 Decodeds | 8499 DomHashs | 1580 Domains |
| 5877 Etags | 778 Favicons | 64 File-Names | 989 HHHashs |
| 434 Images | 198 Ocrs | 144 PGP Dumps | 111 Qrcodes |
| 6895 Titles | 337 User-Accounts | 498 Usernames | |

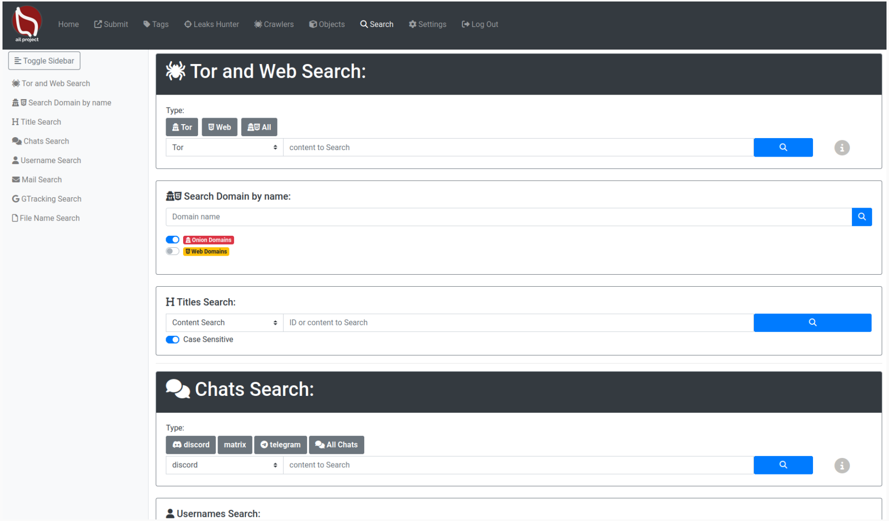
1576 Up
3107 Down
4683 Crawled
14896 Queue
4 UP
0 DOWN
4 Crawled
0 Queue

| | tracker | | Time |
|---|---|---|---|
| yara | Track drug market | ● | 2024-12-07 16:25:49 |
| yara | Track drug market | ● | 2024-12-07 16:25:39 |

| | ID | Tags | Time |
|---|---|---|---|
| ● | crawled/2024/12/07/[redacted]yxwj ywyh44v7ixocqwuelr[redacted]ac2 c6-8339-40a5-937e-e84b7ae351fb | infoleak:submission="crawler" | 16:26:01 |
| ● | | infoleak:automatic-detection="bitcoin-address" | |

TLP:CLEAR

# Search



## Tor and Web Search:

Type:

Tor | content to Search

### Search Domain by name:

Domain name

### Titles Search:

Content Search | ID or content to Search

Case Sensitive

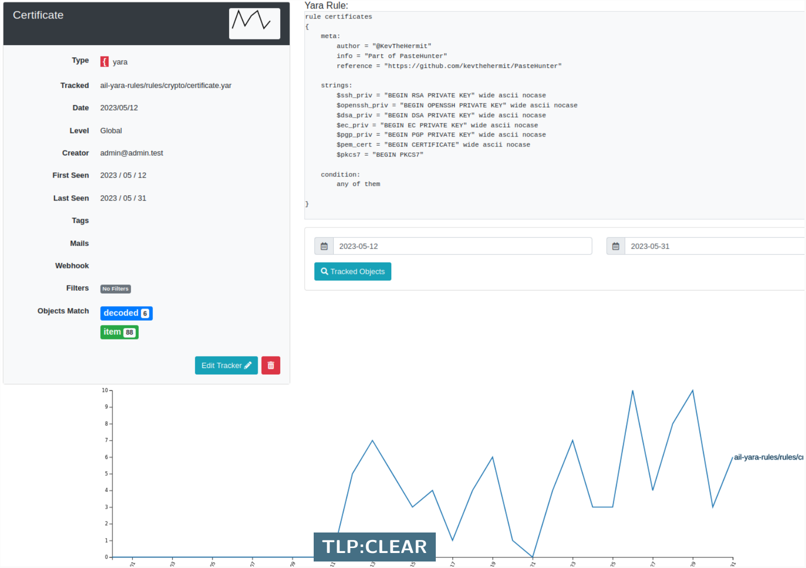
## Chats Search:

Type:

discord | content to Search

### Usernames Search:

# YARA Tracker



Certificate

Type: yara

Tracked: ail-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<br>item 88

Yara Rule:

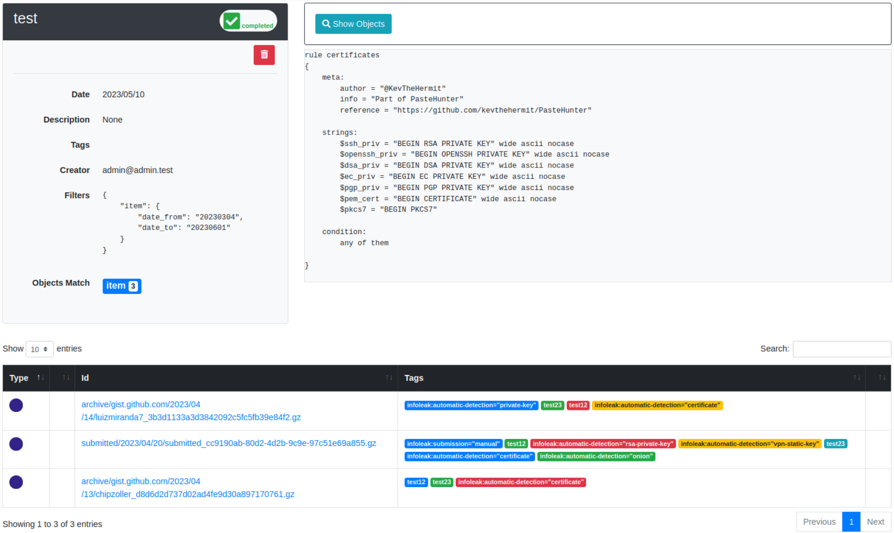
```
rule certificates
{
    meta:
        author = "@KevTheHermit"
        info = "Part of PasteHunter"
        reference = "https://github.com/kevthehermit/PasteHunter"

    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
        $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

}
```

2023-05-12

2023-05-31

TLP:CLEAR

# Retro Hunt



**Date** 2023/05/10

**Description** None

**Tags**

**Creator** admin@admin.test

**Filters**

```
{
    "item": {
        "date_from": "20230304",
        "date_to": "20230601"
    }
}
```

**Objects Match** item 3

```
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

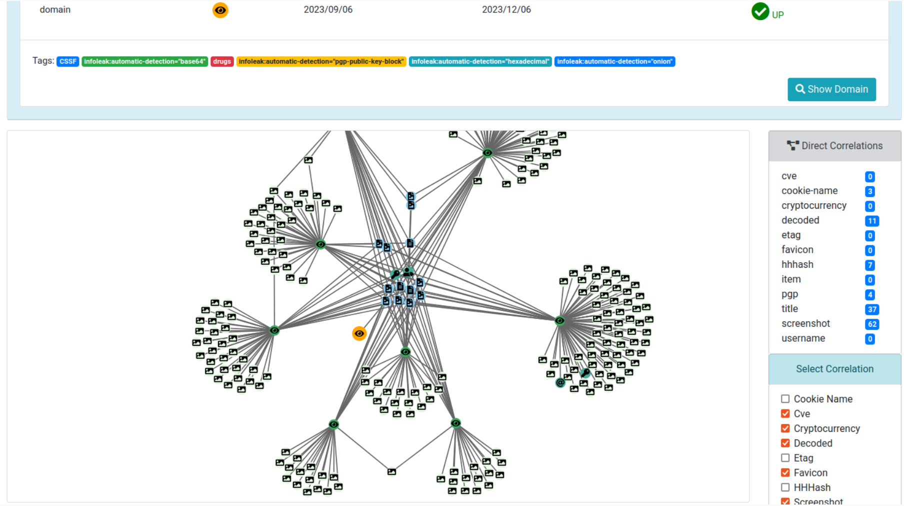
Search:

| 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 1 to 3 of 3 entries

TLP:CLEAR

# Correlations and relationship



| | | | |
|---|---|---|---|
| domain | 2023/09/06 | 2023/12/06 | ✓ UP |

Tags: CSSF infoleak:automatic-detection="base64" drugs infoleak:automatic-detection="pgp-public-key-block" infoleak:automatic-detection="hexadecimal" infoleak:automatic-detection="onion"

Direct Correlations

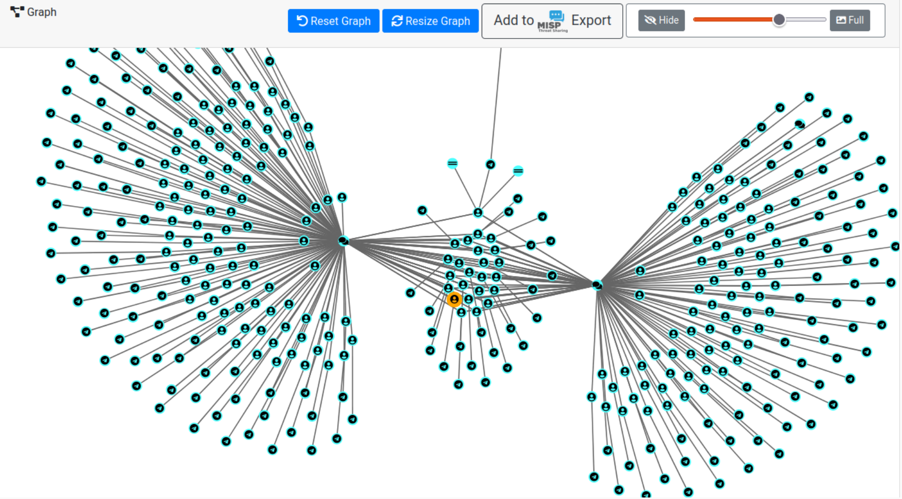
| | |
|---|---|
| cve | 0 |
| cookie-name | 3 |
| cryptocurrency | 0 |
| decoded | 11 |
| etag | 0 |
| favicon | 0 |
| hhhash | 7 |
| item | 0 |
| pgp | 4 |
| title | 37 |
| screenshot | 62 |
| username | 0 |

Select Correlation

- ☐ Cookie Name
- ☑ Cve
- ☑ Cryptocurrency
- ☑ Decoded
- ☐ Etag
- ☑ Favicon
- ☐ HHHash
- ☑ Screenshot

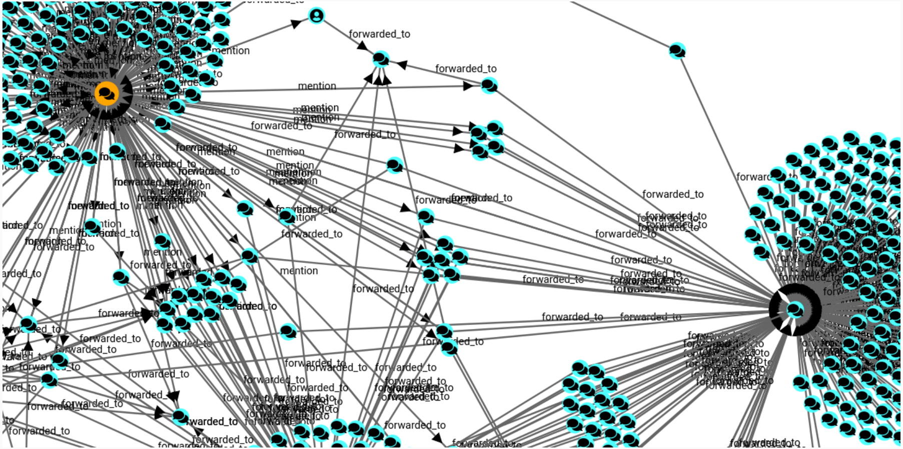
TLP:CLEAR

# User Correlation - Common Chats

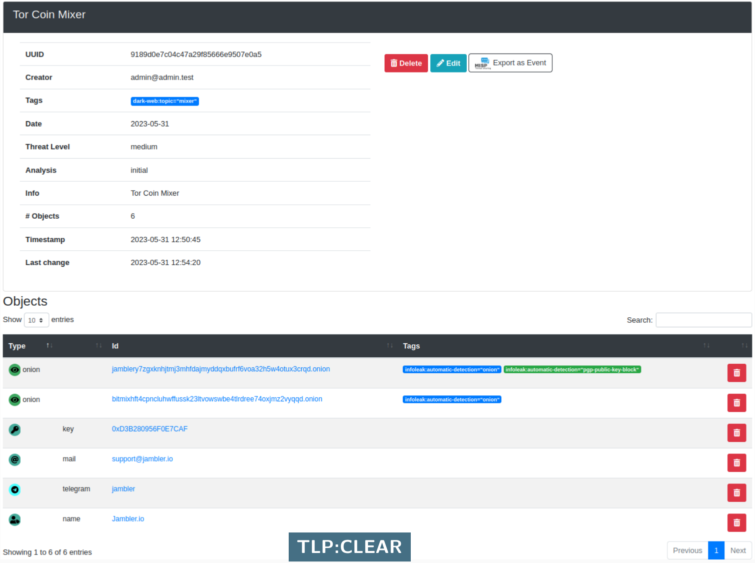


TLP:CLEAR

# What are the Relationships Between Chats Group?



# 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 |

Objects

Show 10 entries

Search:

| Type | | Id | Tags | |
|---|---|---|---|---|
| onion | | jamblery7zgxknhjtmj3mhfdajmyddqxbufrf6voa32h5w4otux3crqd.onion | infoleak:automatic-detection="onion" infoleak:automatic-detection="pgp-public-key-block" | |
| onion | | bitmixhft4cpncluhwffussk23ltvowswbe4tlrdree74oxjmz2vyqqd.onion | infoleak:automatic-detection="onion" | |
| | key | 0xD3B280956F0E7CAF | | |
| | mail | support@jambler.io | | |
| | telegram | jambler | | |
| | name | Jambler.io | | |

Showing 1 to 6 of 6 entries

TLP:CLEAR

# 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[7] : `https://github.com/ail-project/ail-framework`
- For questions, contact: info@circl.lu

[7] 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.