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

SURFnet Workshop

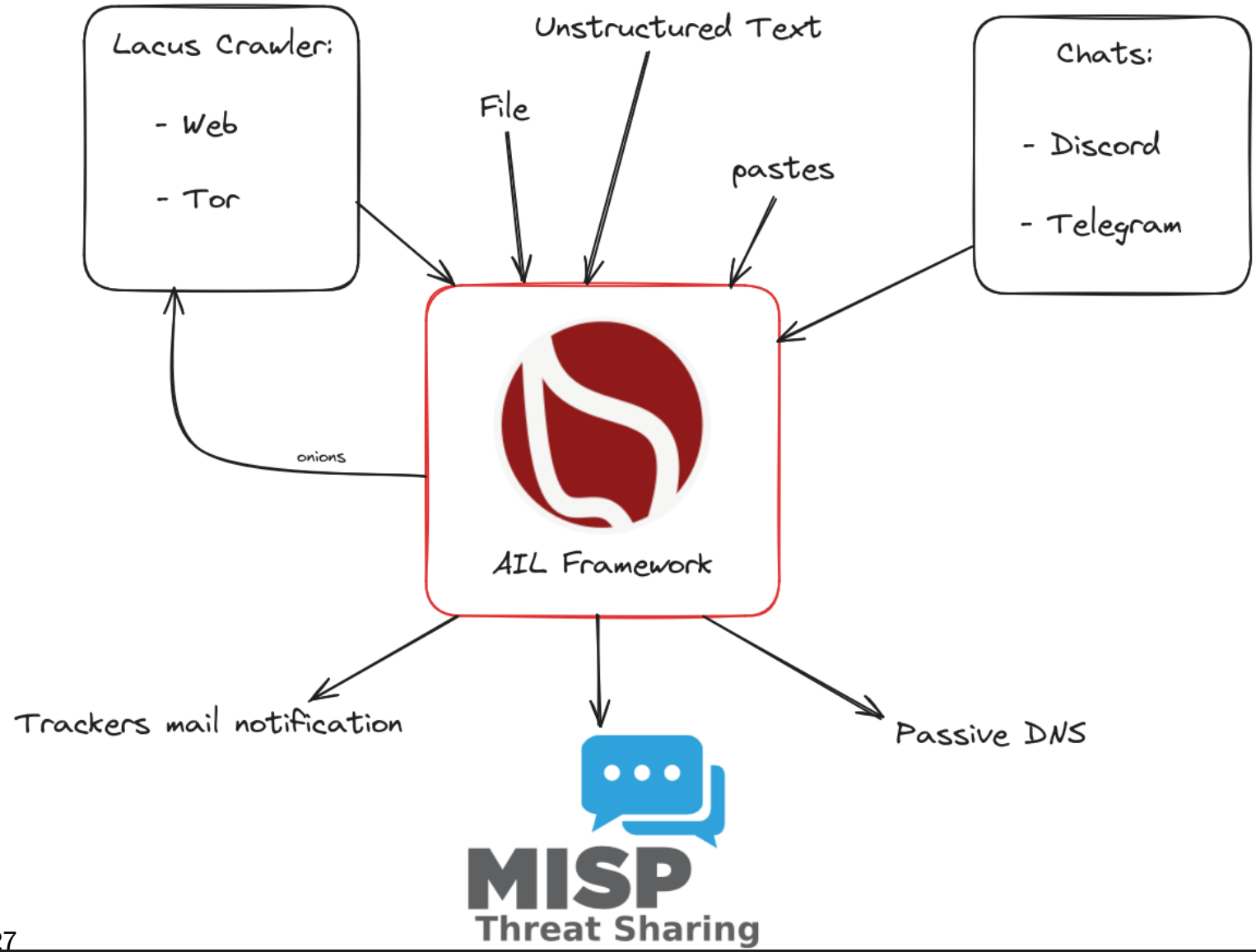
Background

- 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**.

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

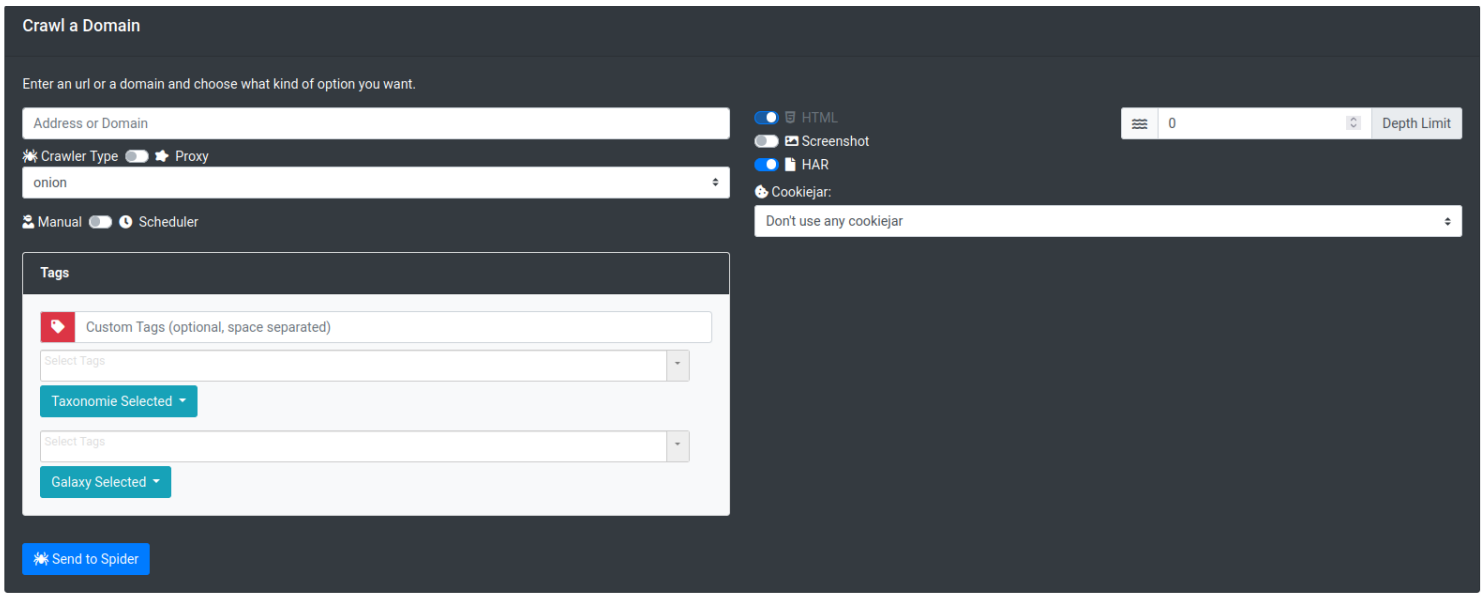
[2] https://github.com/ail-project

High Level Overview



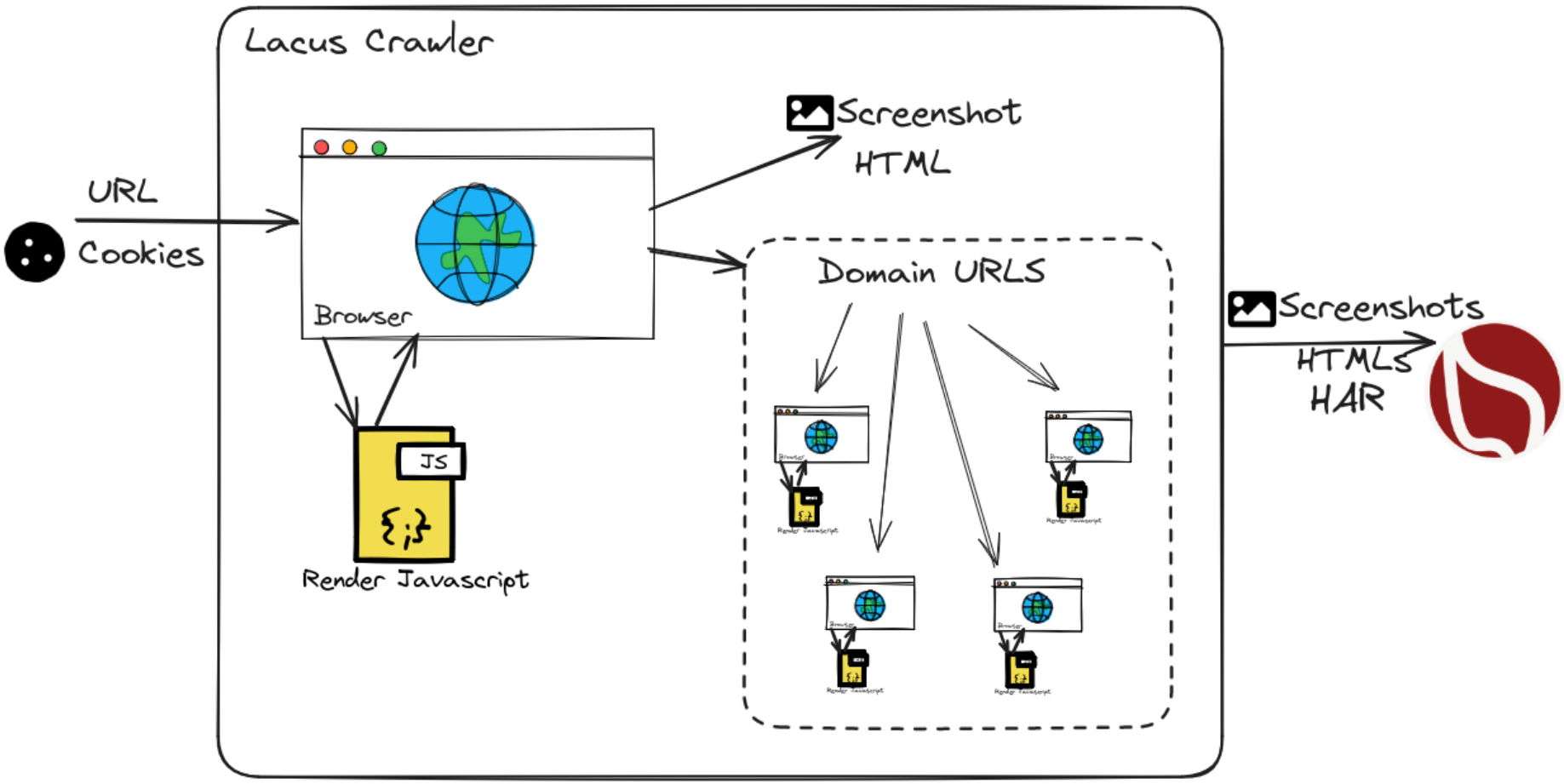
Collection - automate crawling

- 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/`

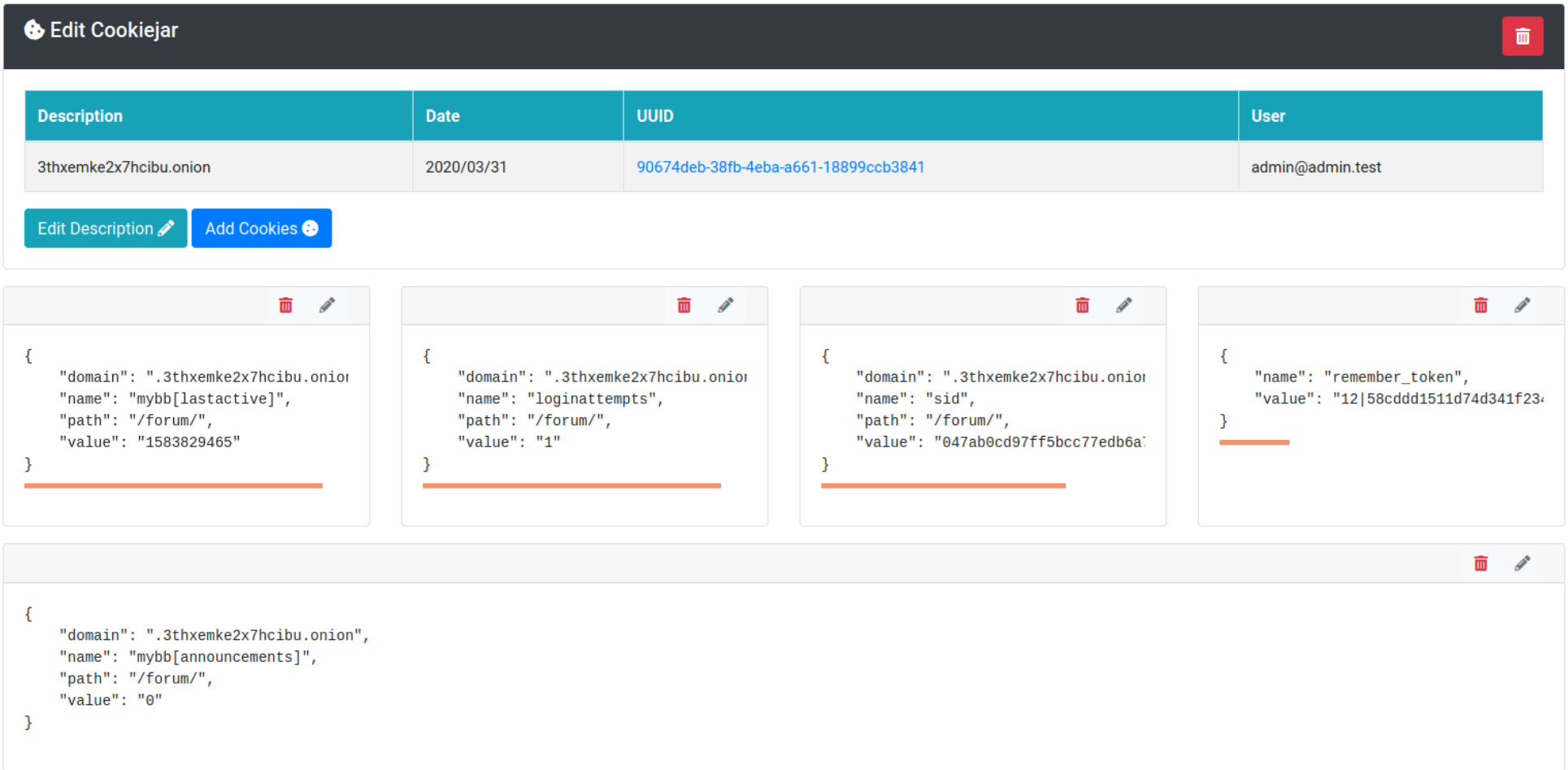
Collection - Lacus Crawler[4]



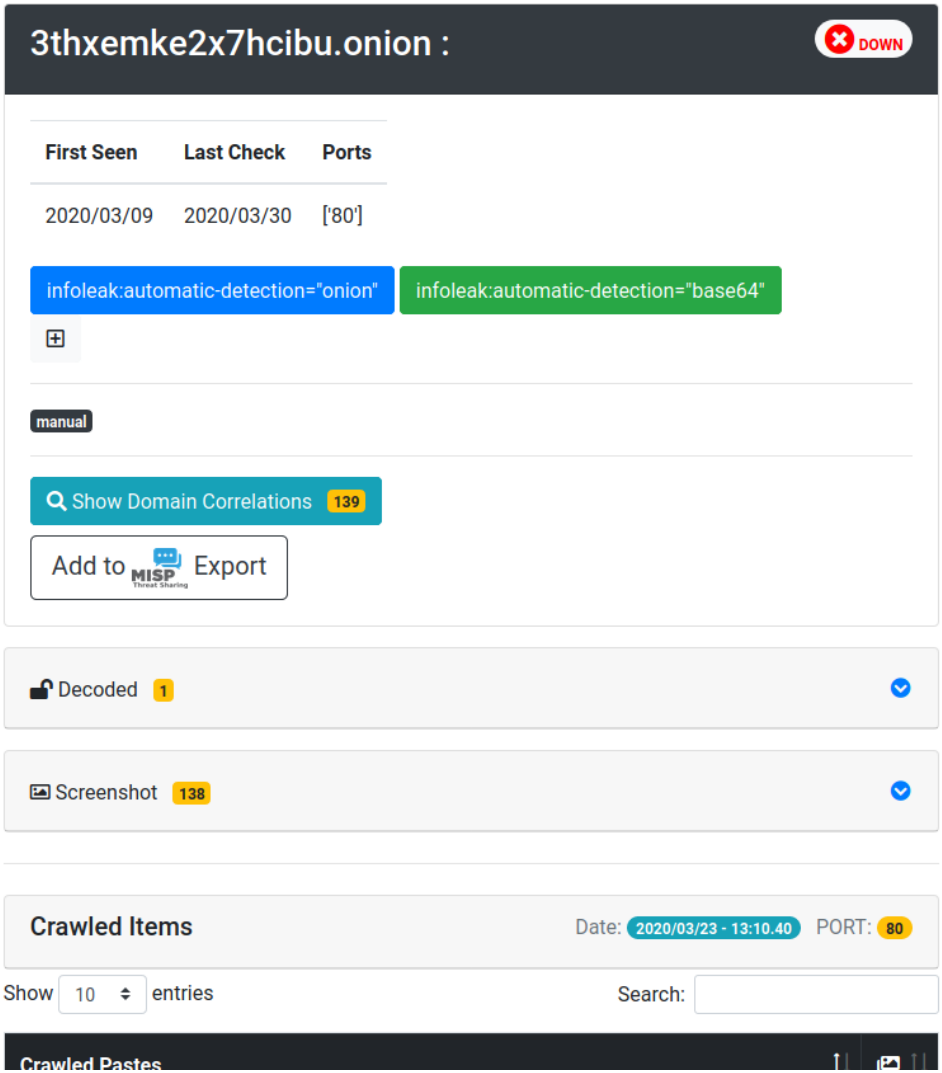
[4] https://github.com/ail-project/lacus

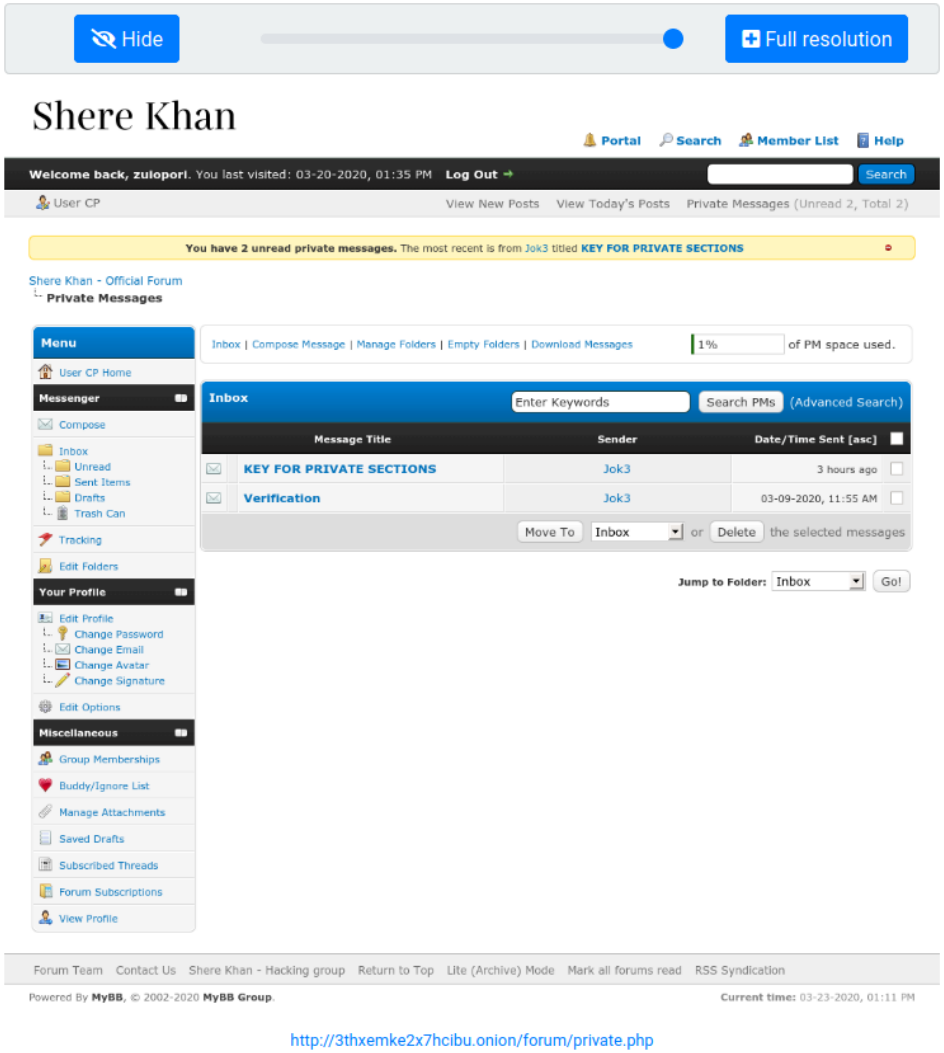
Crawler: Cookiejar

Use your cookies to login and bypass captcha



Crawler: Cookiejar

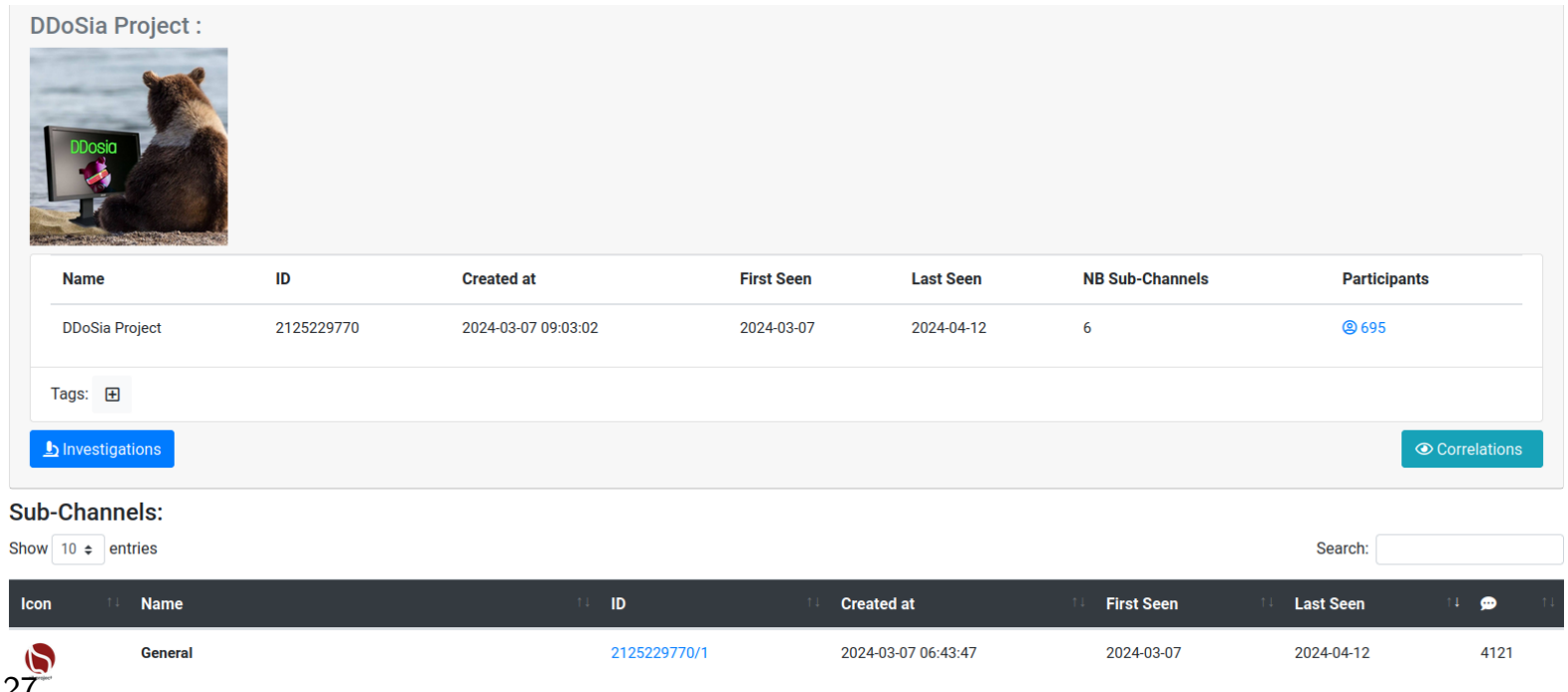




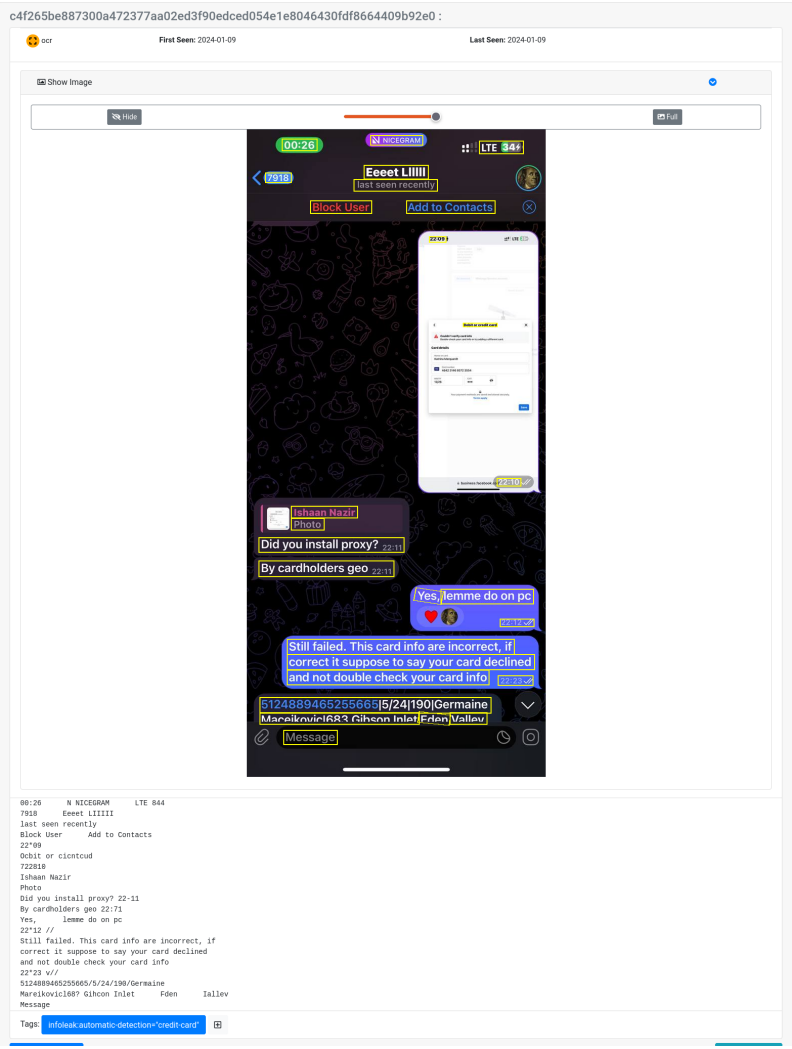
http://3thxemke2x7hcibu.onion/forum/private.php

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.

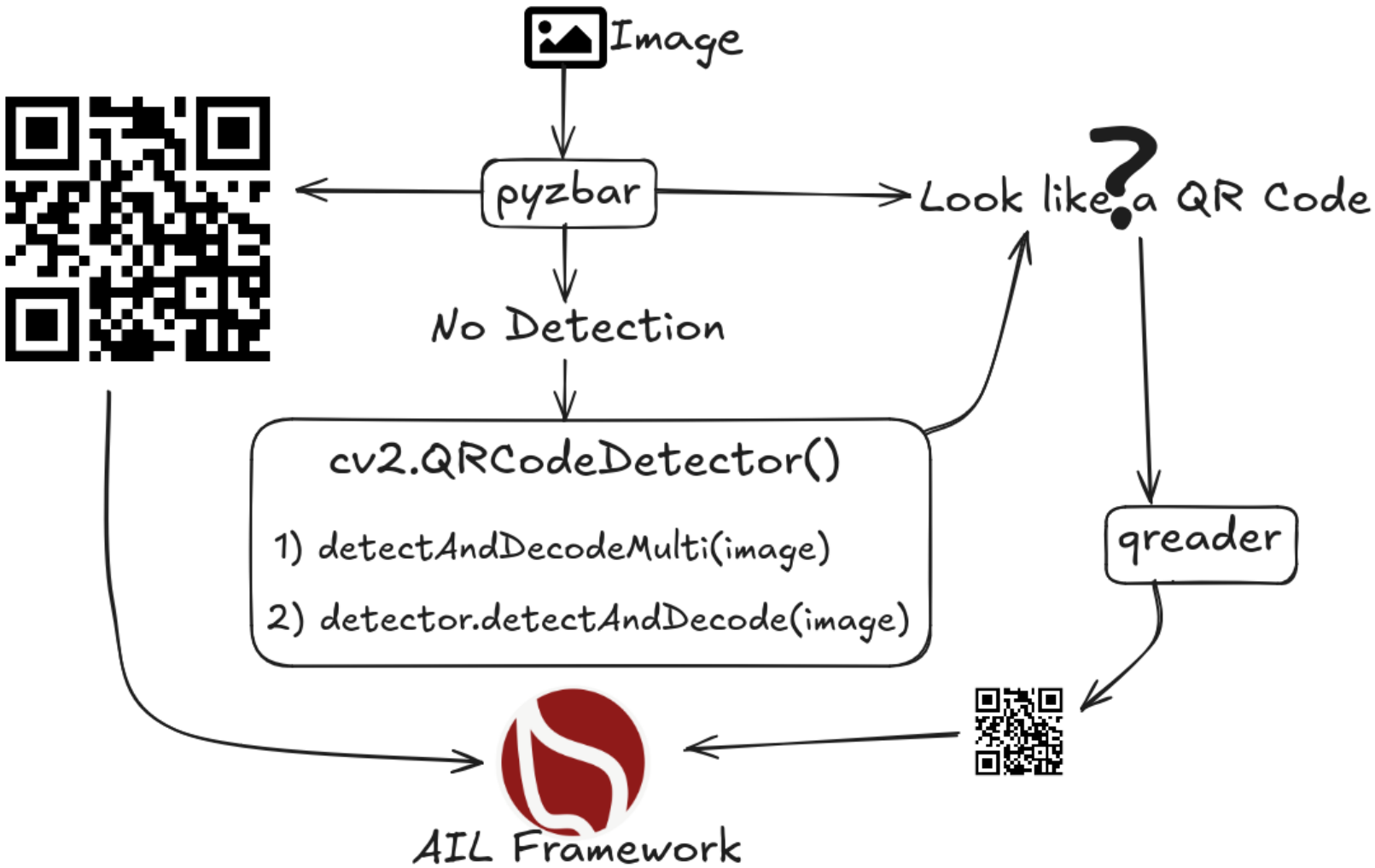


OCR: Optical Character Recognition



- 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**.

QR Code Extractor



QR Codes

2022-12-16
16:51:04

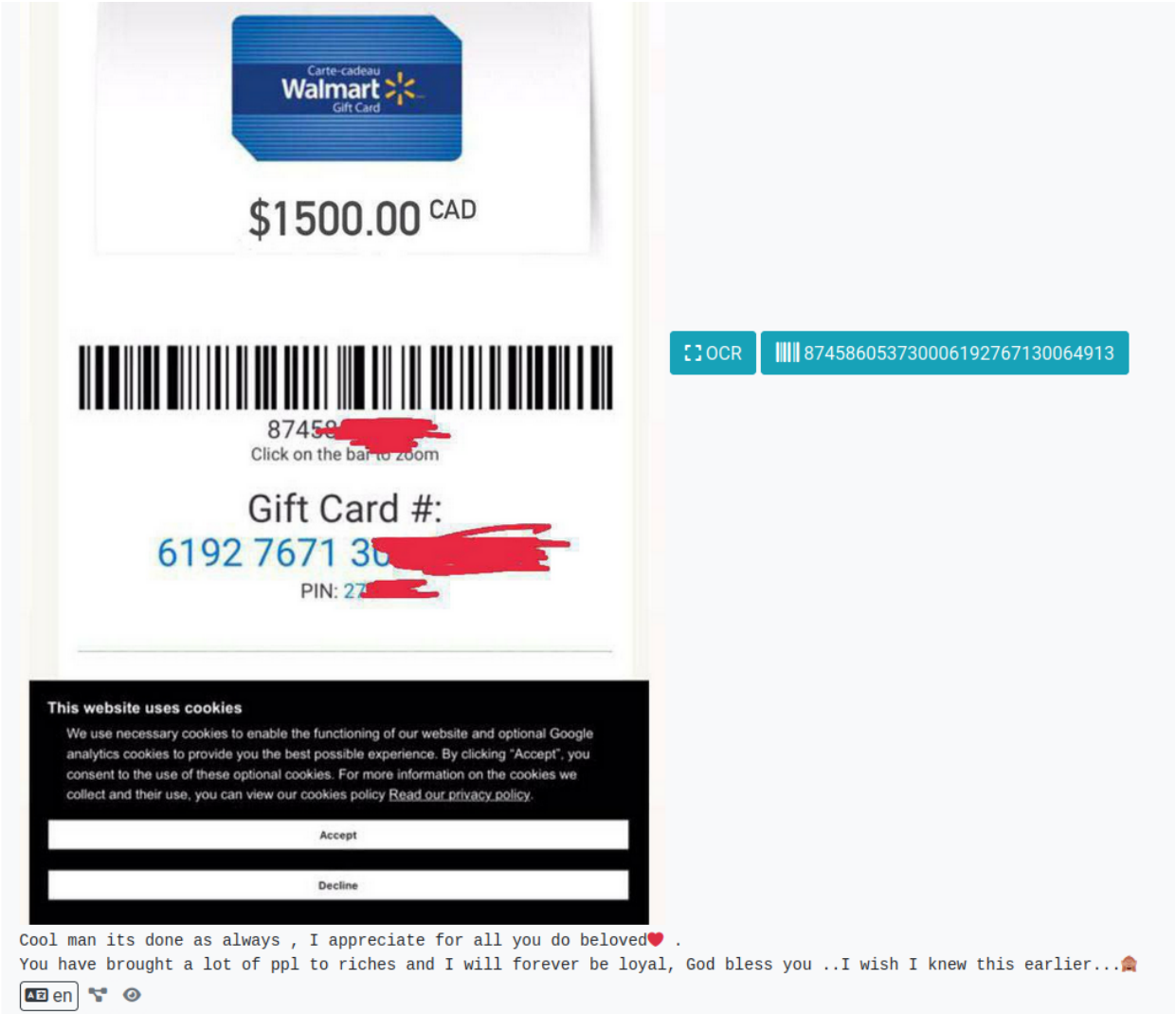
blue_forever20

OCR

0002010102110216433273000096494304155290090000964941531267607642676076400
Mai62360520AP9J3I0000000000000007084475831663O4F69D

Someone with union pay link or QR code
Hit me up

Bar Codes


Carte-cadeau
Walmart
Gift Card
$1500.00 CAD
Click on the bar to zoom
Gift Card #:
PIN:
OCR
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.
Accept
Decline
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...
en

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`

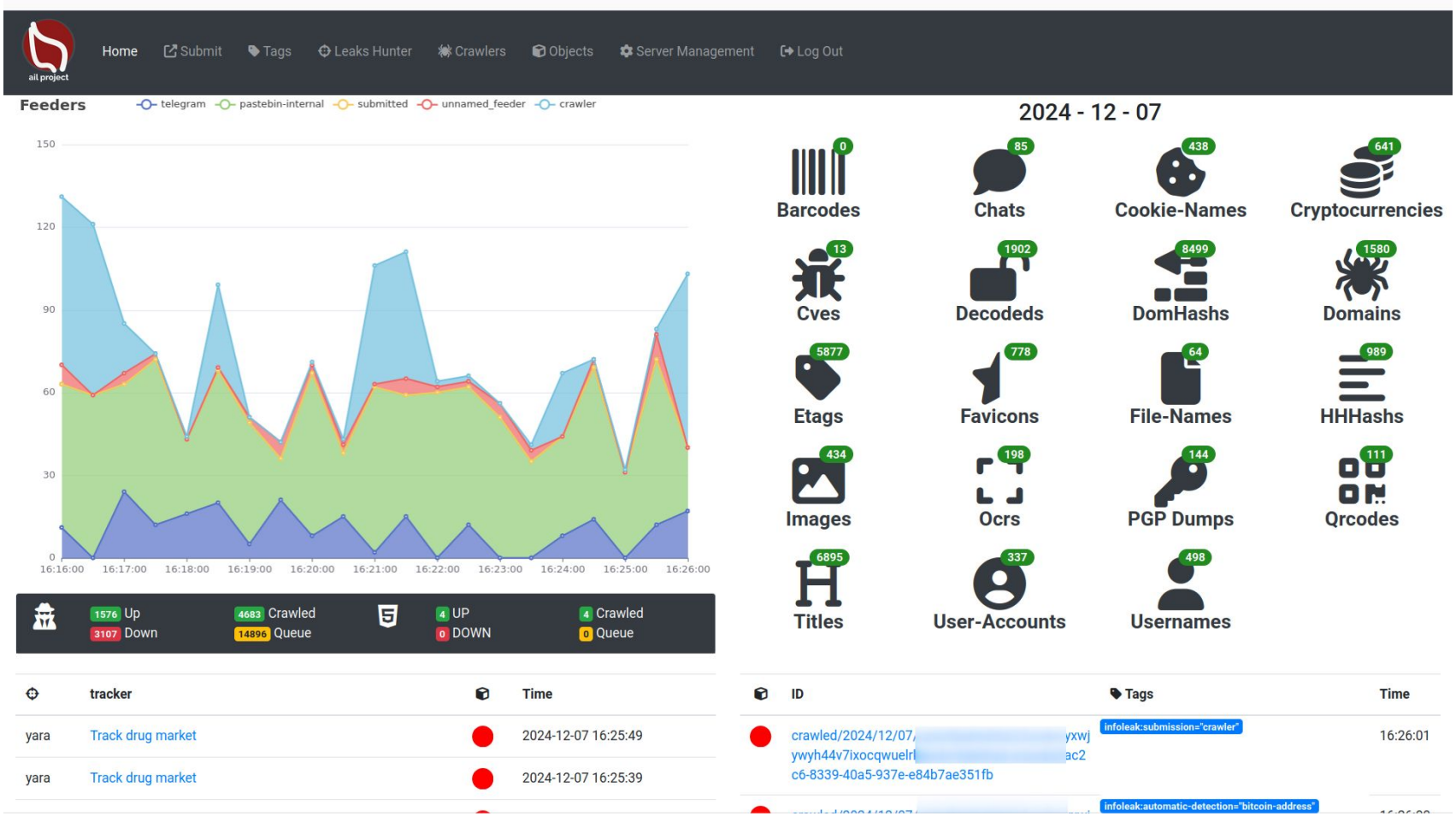
[6] Relying on MISP taxonomies and galaxy

Live Trackers & Retro Hunt

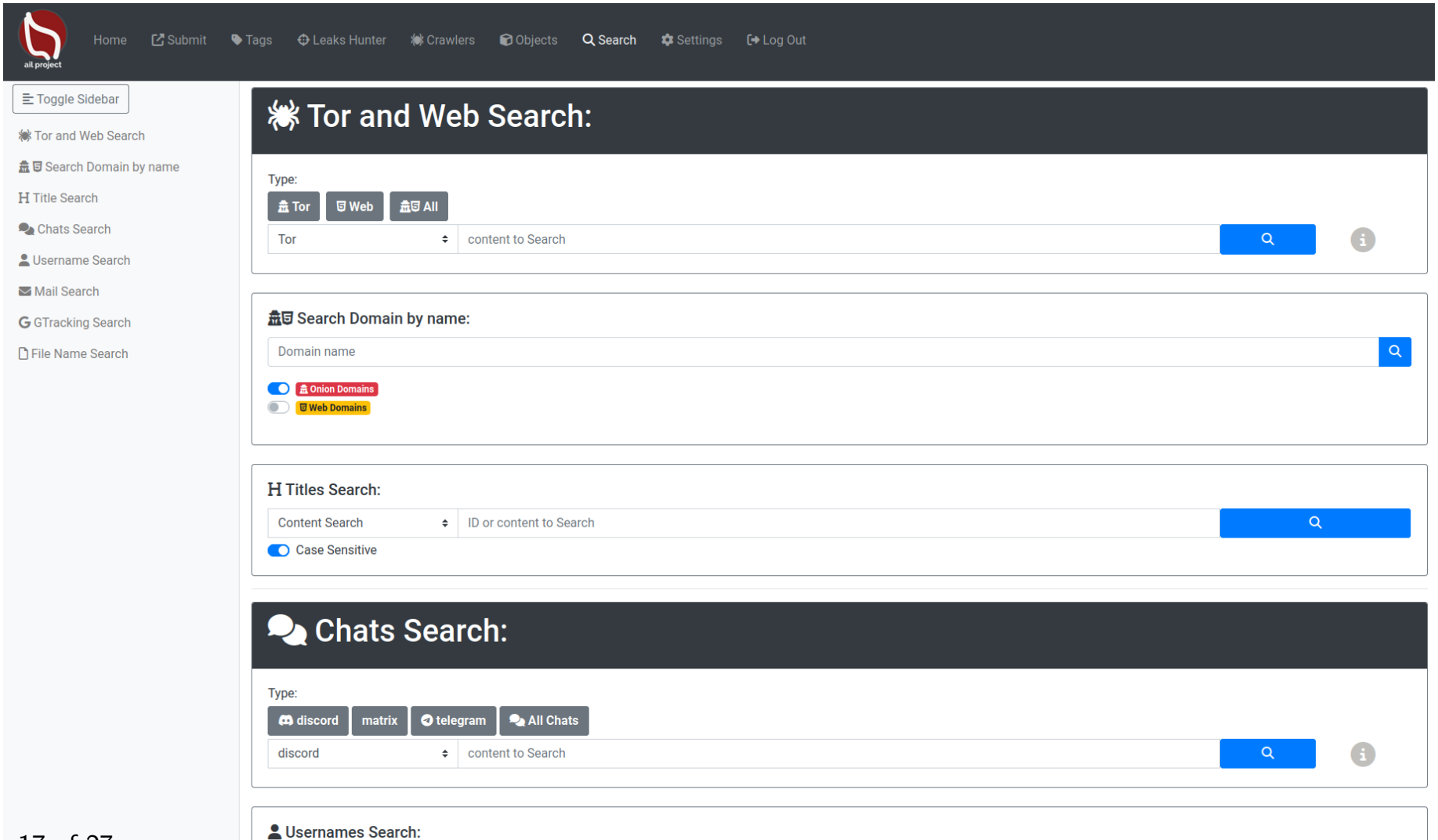
- Search and monitor specific keywords/patterns
  - Automatic tagging
  - Email/webhook notifications
- Track Word
  - ddos
- Track Set
  - booter, ddos, stresser; 2
- Track Regex
  - circl\lu
- **YARA rules**
  - `https://github.com/ail-project/ail-yara-rules`

Live Demo

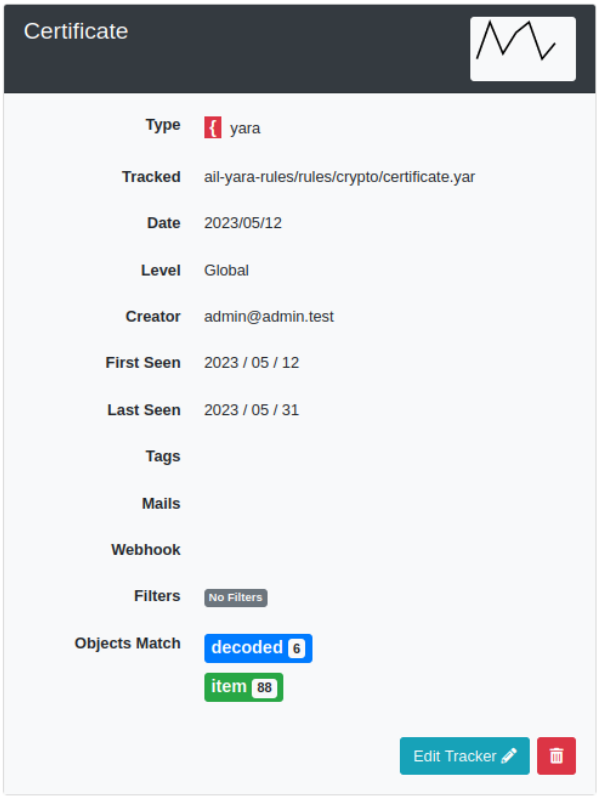
Dashboard


Home
Submit
Tags
Leaks Hunter
Crawlers
Objects
Server Management
Log Out
Feeders
telegram
pastebin-internal
submitted
unnamed_feeder
crawler
2024 - 12 - 07
Barcodes
Chats
Cookie-Names
Cryptocurrencies
Cves
Decodeds
DomHashs
Domains
Etags
Favicons
File-Names
HHHashs
Images
Ocrs
PGP Dumps
Qrcodes
Titles
User-Accounts
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
infoleak:submission="crawler"
16:26:01
infoleak:automatic-detection="bitcoin-address"

Search


all project
Home
Submit
Tags
Leaks Hunter
Crawlers
Objects
Search
Settings
Log Out
Toggle Sidebar
Tor and Web Search
Search Domain by name
Title Search
Chats Search
Username Search
Mail Search
GTracking Search
File Name Search
Tor and Web Search:
Type:
Tor
Web
All
Tor
content to Search
Search Domain by name:
Domain name
Onion Domains
Web Domains
Titles Search:
Content Search
ID or content to Search
Case Sensitive
Chats Search:
Type:
discord
matrix
telegram
All Chats
discord
content to Search
Usernames Search:

YARA Tracker



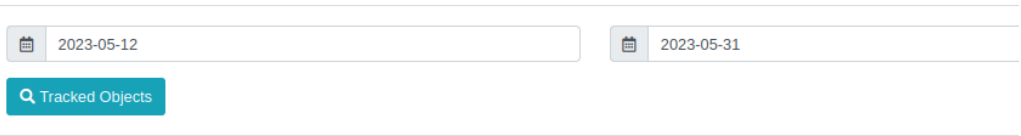
Yara Rule:

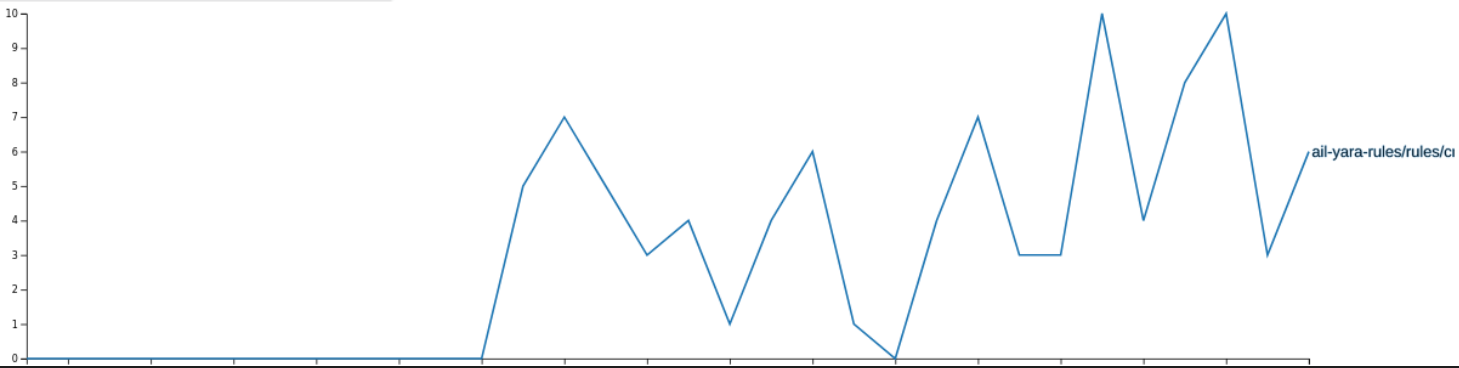
```
rule certificates
{
    meta:
        author = "@KevTheHermit"
        info = "Part of PasteHunter"
        reference = "https://github.com/kevthehermit/PasteHunter"

    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
        $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

}
```





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

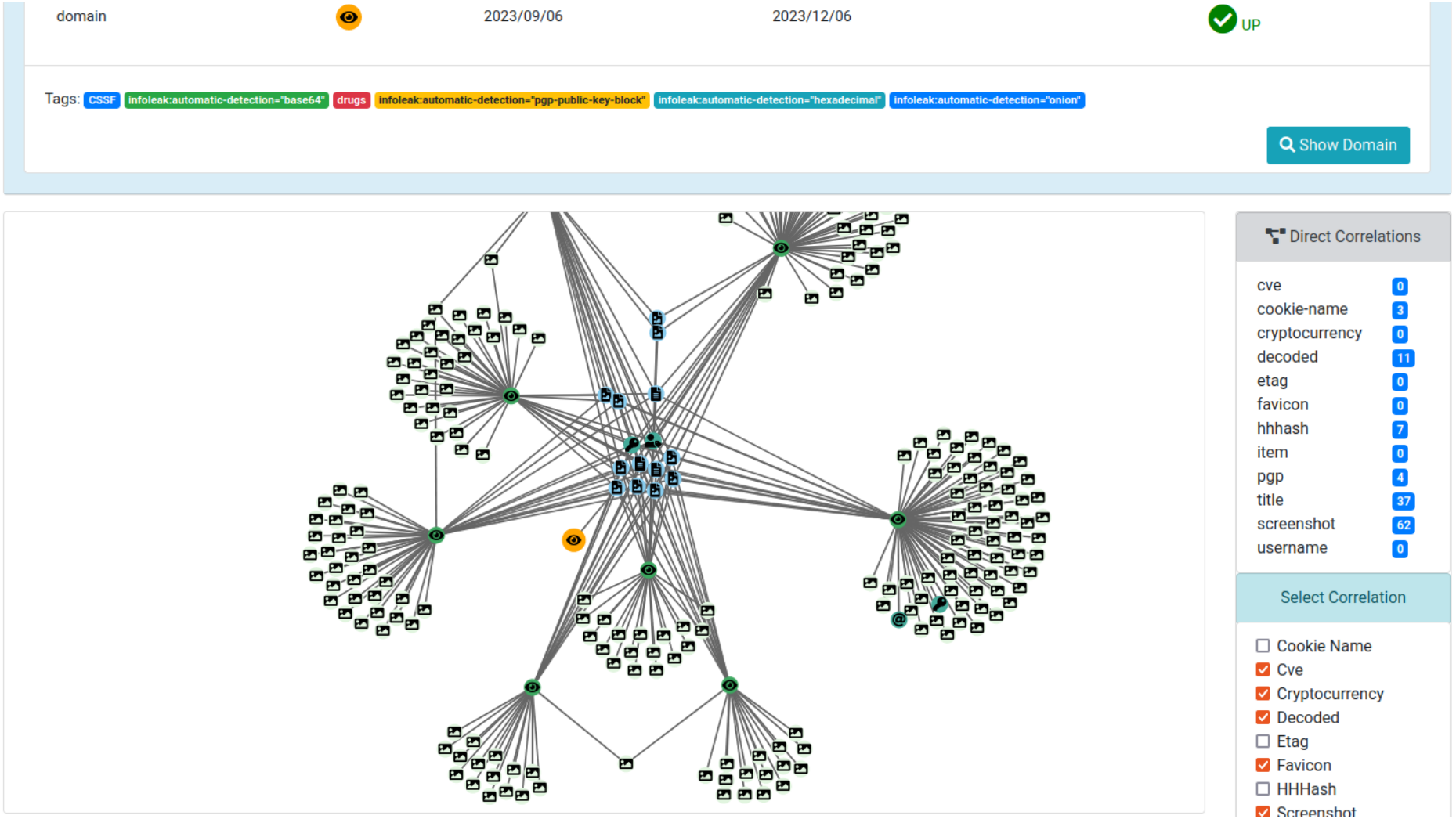
Search:

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"	

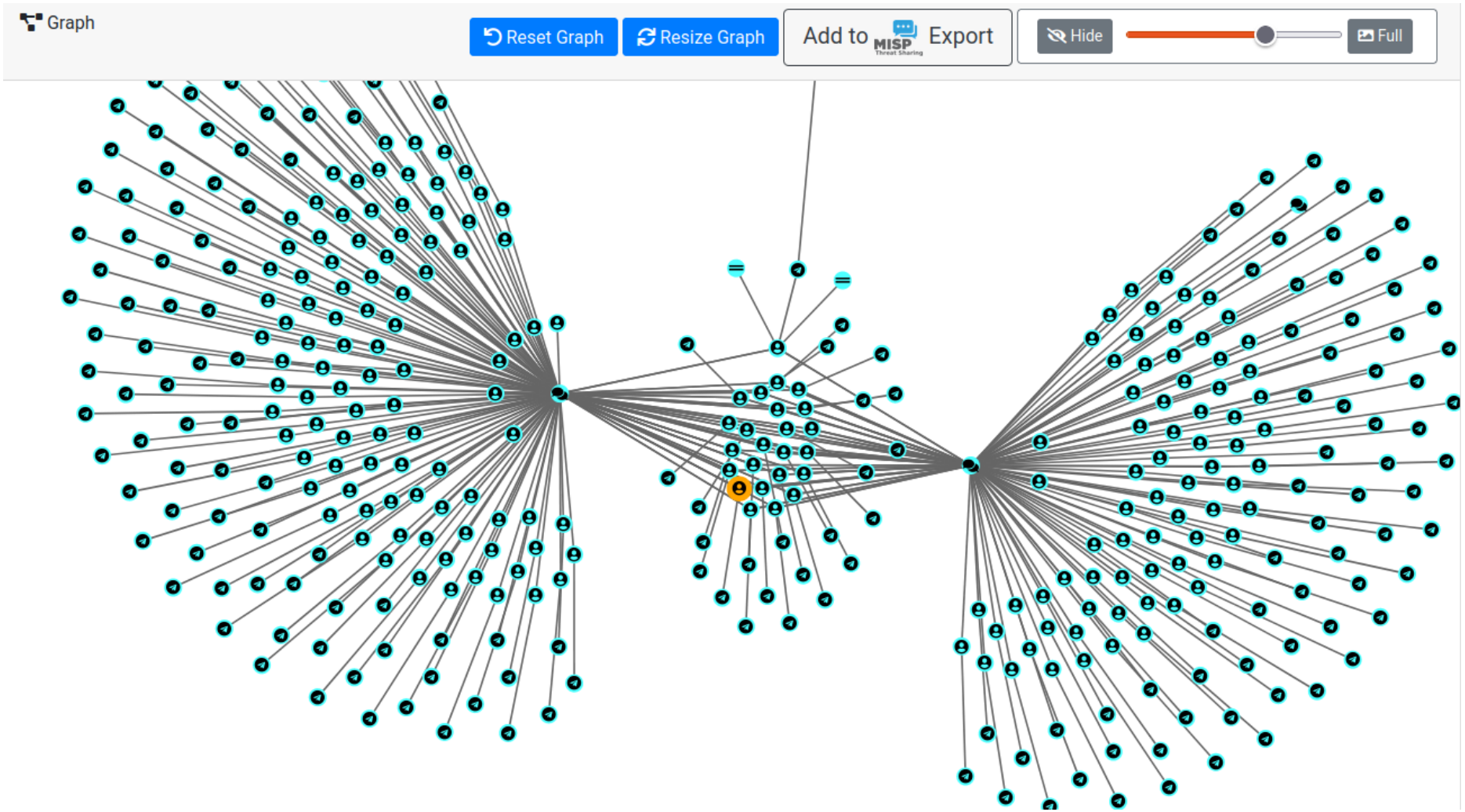
Showing 1 to 3 of 3 entries

Previous 1 Next

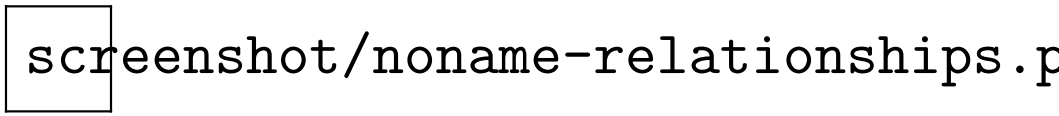
Correlations and relationship



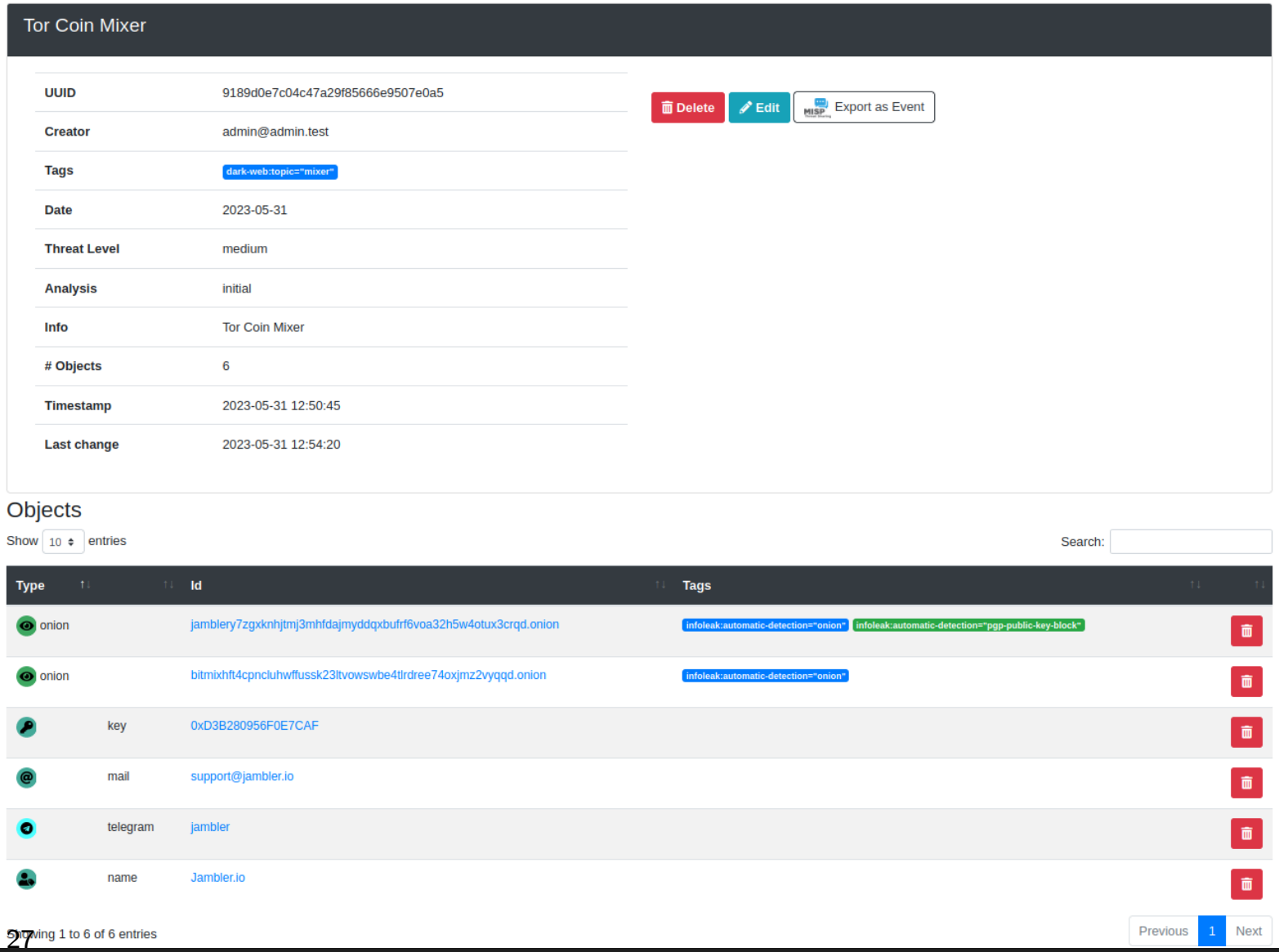
User Correlation - Common Chats



What are the Relationships Between Chats Group?



Investigations

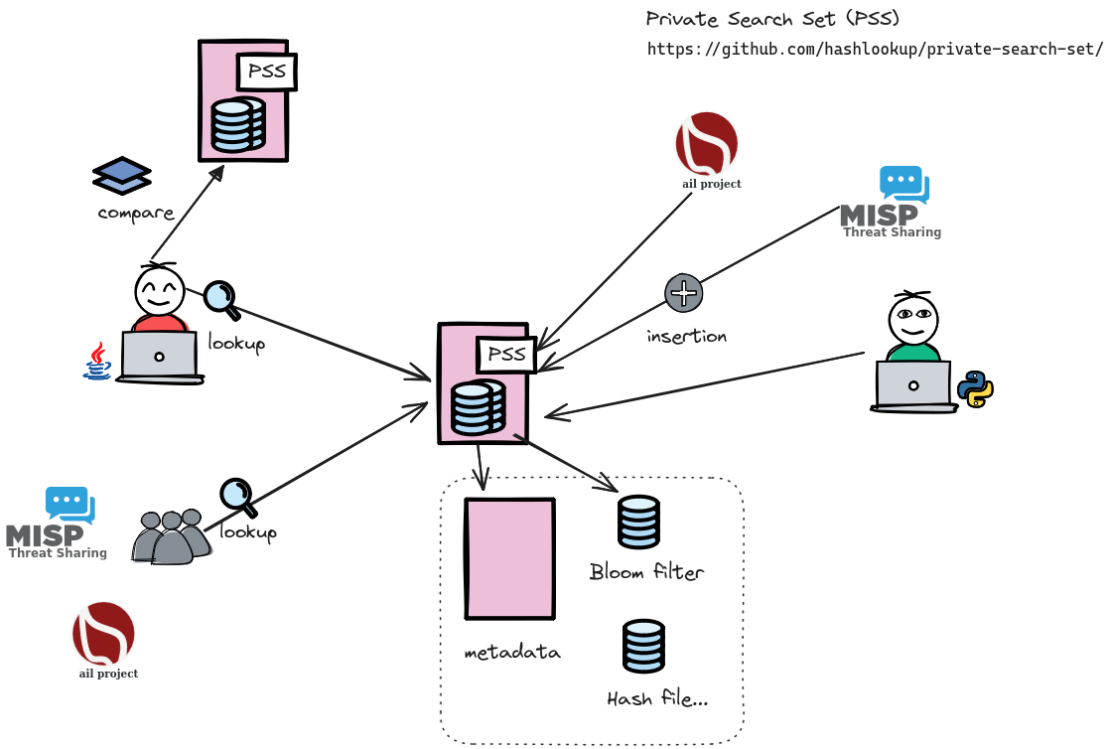


AIL Framework: Extensible Capabilities

- 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

- **Advanced video processing and extraction**
- Improve Search
- **MISP export with new correlation types**
- **Automatic geolocation**
- **Bloom Filter for Private Search Set (PSS) Filtering**



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`

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**