



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

Aurelien Thirion - aurelien.thirion@circl.lu

July 17, 2025

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

Background

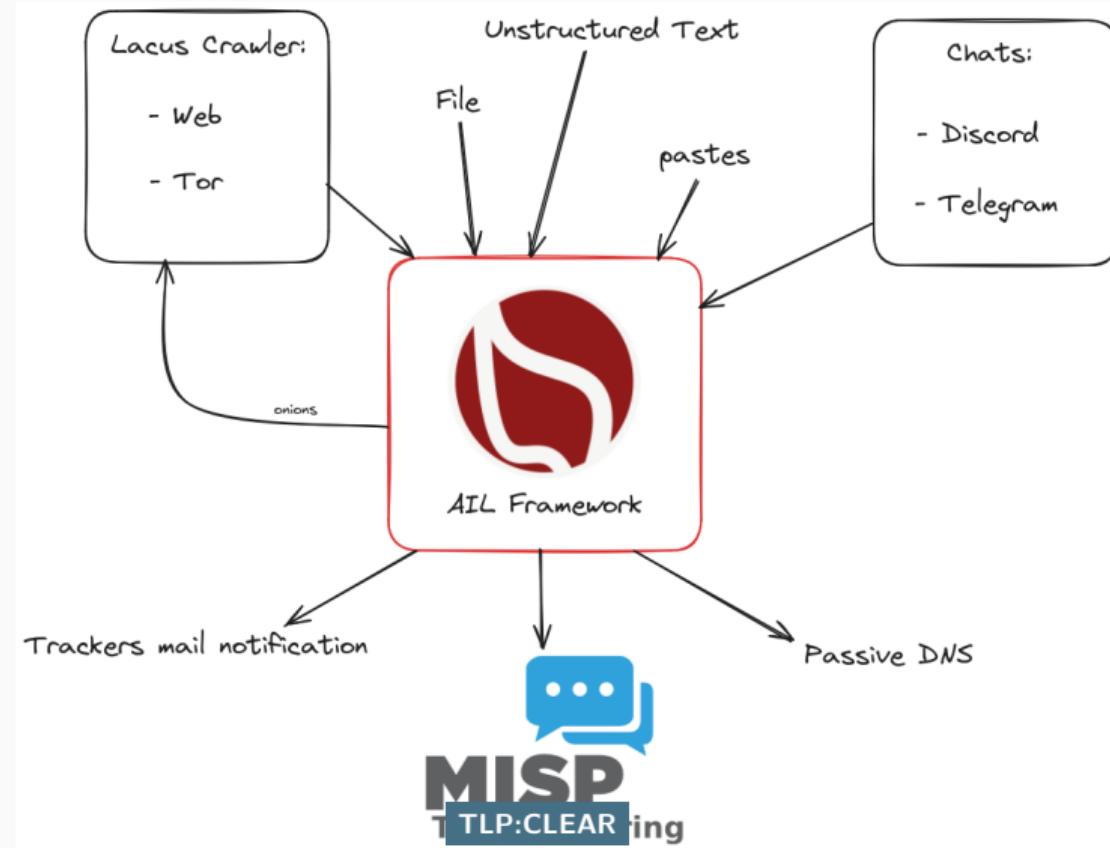
- Over the past years, CIRCL has developed the AIL project¹ to fulfill our needs at CIRCL in intelligence gathering and analysis.
- AIL features an extensible Python-based framework for the **analysis of unstructured information**, collected either through an advanced crawler manager or from various feeders, including social networks and custom feeders.
- The AIL Project is an **open-source** framework² comprising various modules designed for the **collection, crawling, digging, and analysis of unstructured data**.



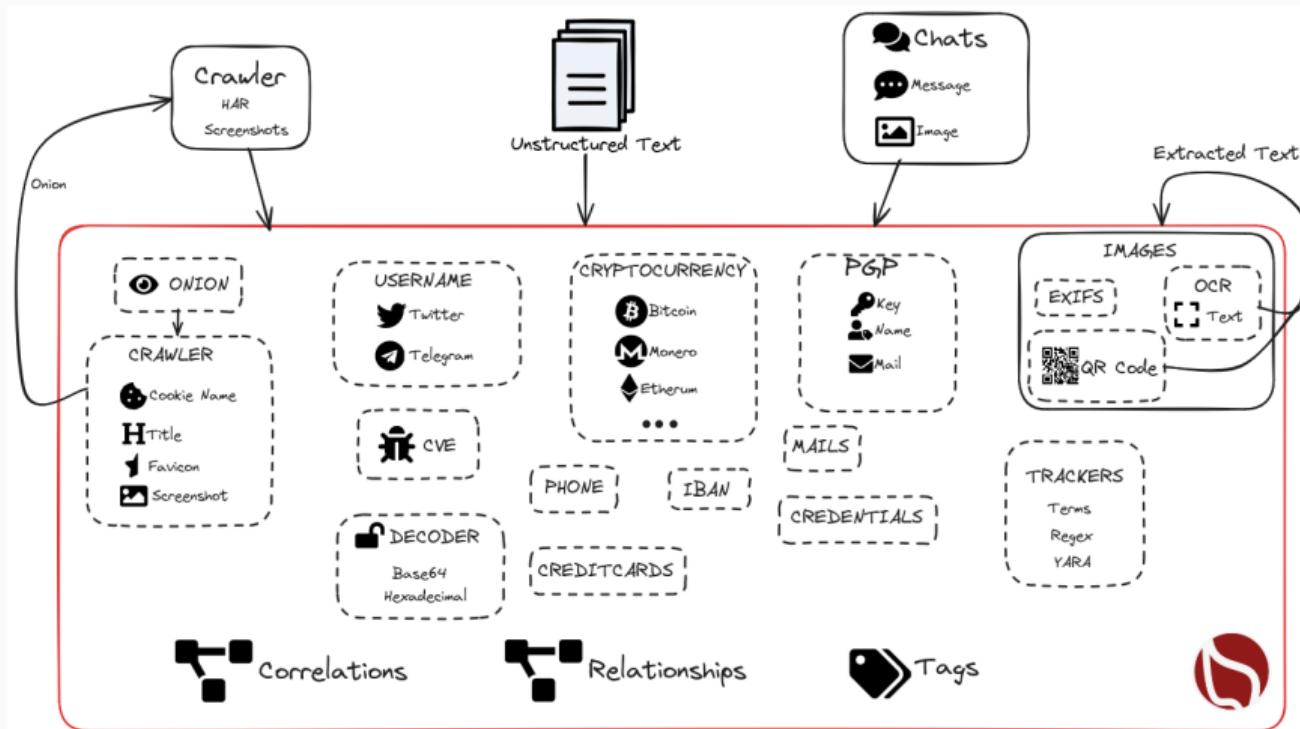
¹<https://www.ail-project.org/>

²<https://github.com/ail-project>

High Level Overview

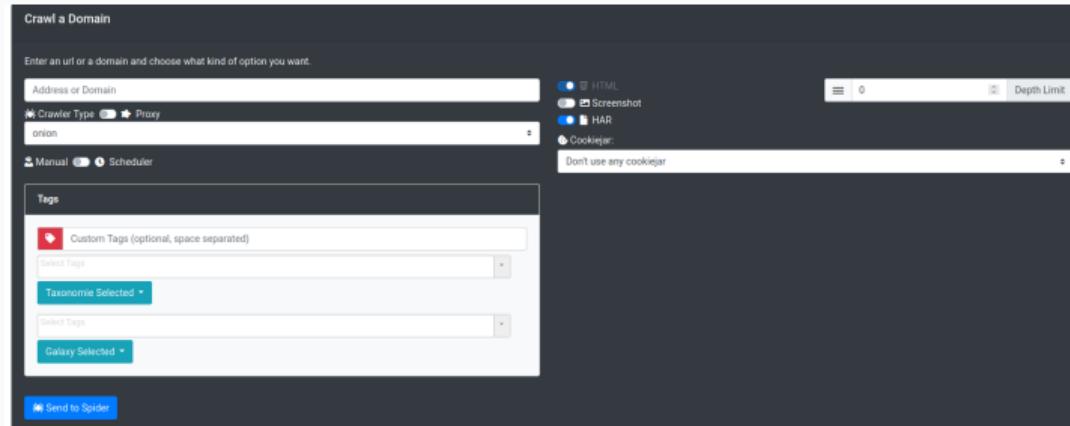


Analysis of unstructured information



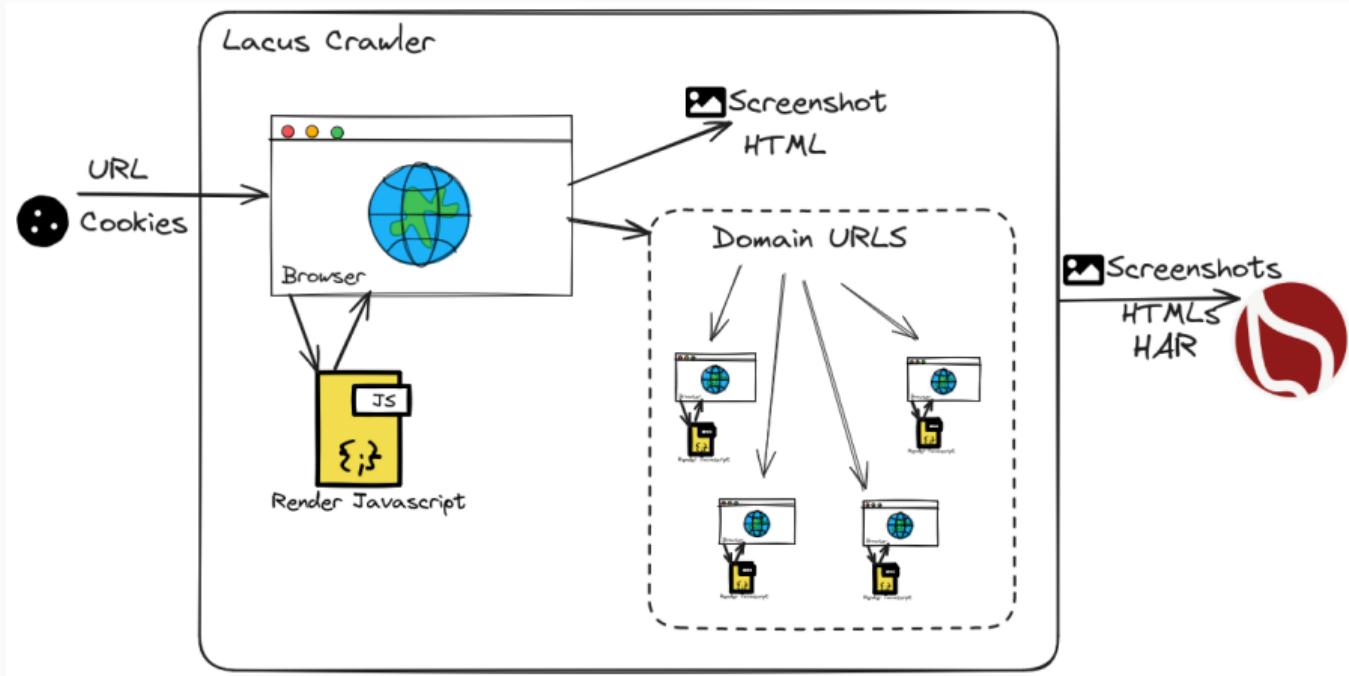
Collection - automate crawling

- Crawling can be a challenging task, for example, gathering all the blog posts from ransomware groups³, 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 Crawler⁴



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](#)

[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" }
```

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 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 BuddyIgnore 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]
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

8/31

Onion Lookup - onion.ail-project.org

- <https://onion.ail-project.org> is a public search engine powered by AIL.
- Check if a Tor hidden service exists and retrieve its associated metadata.
- Can be used to filter onion domains related to CSAM and other violent content.



onion-lookup

onion-lookup is a service for checking the existence of Tor hidden services and retrieving their associated metadata. onion-lookup relies on a private AIL instance to obtain the metadata.

archiveiya74codqgiixo33q62qlrqtkgmcitqx5u2oeqn5bpcliyd.onion Lookup

hqer7pfhpq5wyufhmtp3lseffwiddbctqby262zup4dstab2pjtmqd.onion

① [Infoleak:automatic-detection='onion'](#) [Infoleak:automatic-detection='base64'](#)

Languages ① : en

Known titles of the Tor onion addresses ① :

HQER - High Quality Euro Counterfeits - best counterfeit bank notes in europe

Temporal information ① :

First seen: 2023-04-08

Last seen: 2025-07-04

TLP:CLEAR

Crawler - Pre-Filtering

Filter Unsafe Onion: **False**

This option enables filtering of onion domains that are considered unsafe due to containing violent content, child sexual abuse material (CSAM), or other harmful materials. When enabled, the system will attempt to identify and exclude such domains from crawling.

 **Disabling this option may result in crawling and downloading content that includes CSAM, extreme violence, or other harmful materials.**

Users are strongly advised to keep this feature enabled to avoid unintentional exposure to such content.

 **How It Works:** The filtering mechanism leverages known blocklists, heuristics, and automated detection techniques to reduce the risk of crawling unsafe content. While no filtering system is perfect, we continuously strive to improve detection and minimize exposure to harmful materials.

By using this feature, you benefit from an added layer of protection, but please note that some unsafe onion domains may still bypass detection due to evolving content and obfuscation techniques. We encourage users to remain cautious and use this feature as an additional safeguard.

Enable Onion Filter

Crawl Unknown Onion: **Disabled**

This option controls whether the crawler should proceed with crawling onion domains that have **not yet been classified** as safe or unsafe.

- **If disabled:** The crawler will process domains that have never been checked, potentially discovering new useful content but also increasing the risk of encountering unsafe materials.
- **If enabled:** The crawler will only process domains that have been explicitly identified as safe, reducing risk but potentially missing new, unclassified domains.

This option is useful for users who want to explore uncharted domains while still benefiting from the `filter_unsafe_onion` protection. However, enabling this option increases the likelihood of encountering harmful content, so caution is advised.

Enable Unknown Onion Filter

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: <#>

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

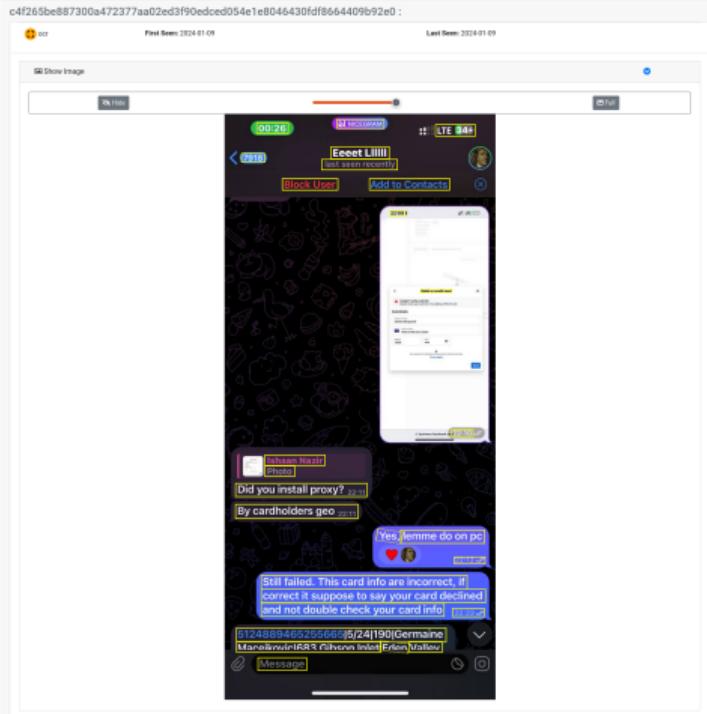
Sub-Channels:

Show [10](#) entries

Search:

Icon	Name	ID	Created at	First Seen	Last Seen	Participants
	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

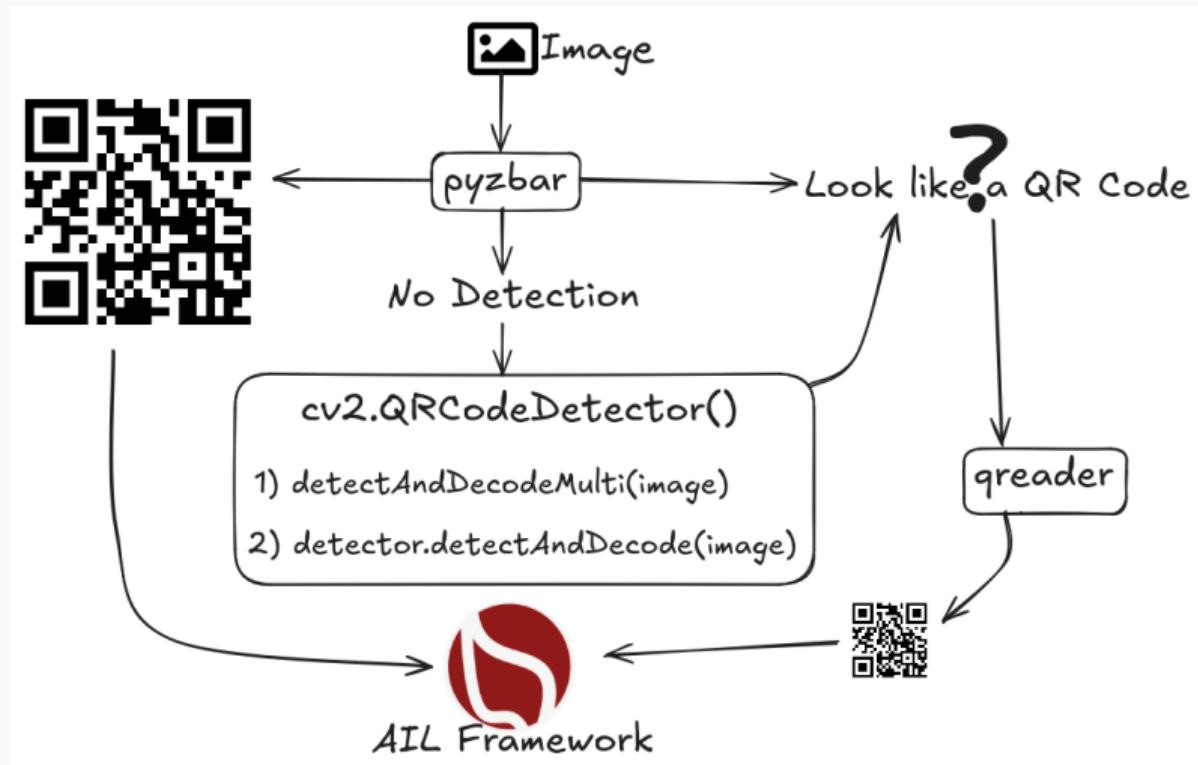
OCR: Optical Character Recognition



- 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



2022-12-16
16:51:04

blue_forever20



OCR

likeaboss

000201010211021643327300096494304155290090009649415312676076426760764000
Mai62360520AP9J3I00000000000000000708447583166304F69D

Someone with union pay link or QR code

Hit me up

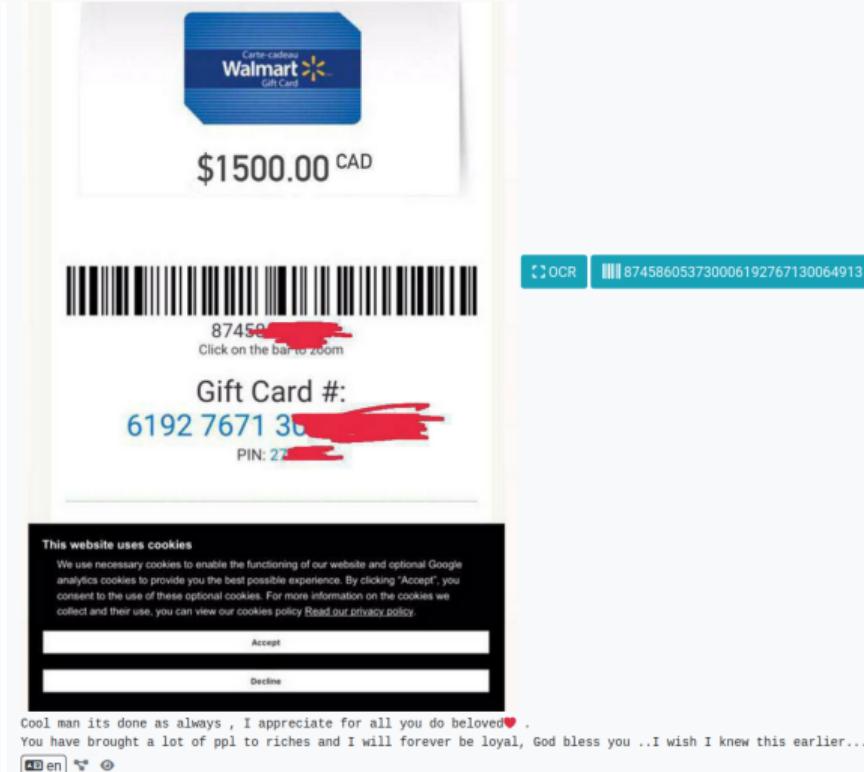
TLP:CLEAR

We can make money

en

•

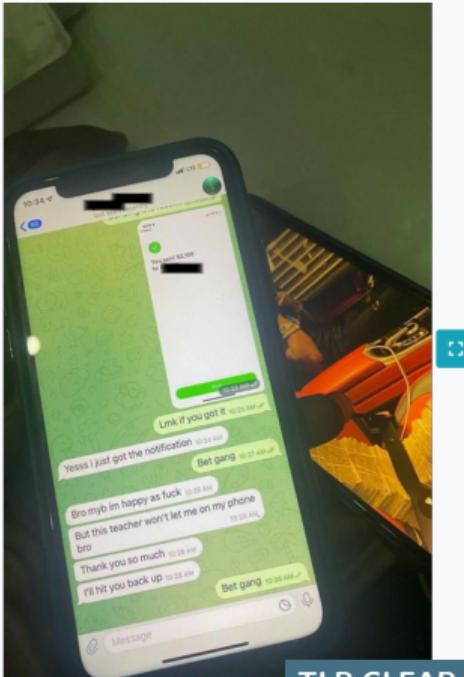
Bar Codes



TLP:CLEAR

Images Descriptions

- Uses **Ollama** and **Qwen2.5-VL** to automatically generate descriptions for screenshots and other images.



TLP:CLEAR

The picture shows a smartphone displaying a WhatsApp conversation. The conversation is between two people, with one person sending a message and the other responding. The message content appears to be about a financial transaction, with the sender mentioning sending \$2.100 to someone named ████. The conversation also includes some casual and informal language, with phrases like "Lmk if you got it," "Yesss i just got the notification," and "Bro myb im happy as fuck." The time stamps on the messages indicate that the conversation took place around 10:23 AM to 10:28 AM. The background of the phone screen shows a wallpaper with cartoonish characters and a green theme.

Images Descriptions

- Helps users quickly understand image content without viewing the image directly.
- Generated descriptions provide an extra layer of insight during analysis.



The image shows a black tactical scope mount. This type of mount is typically used for attaching a scope to a firearm, such as a rifle. The mount has a cylindrical opening in the center, which is designed to accept a scope. The mount also has several mounting points and adjustment screws, allowing for precise alignment and adjustment of the scope. The mount appears to be made of metal and is designed for durability and stability. The background shows a wooden surface with some packaging materials, indicating that the mount is new and possibly still in its original packaging.

- Extracting **credit card numbers, credentials, phone numbers, ...**
- Extracting and validating potential **hostnames**
- Submission to threat sharing and incident response platforms (**MISP** and **FlowIntel⁵**)
- **Tagging⁶** 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>

⁶Relying 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

2024 - 12 - 07

Category	Count
Barcodes	1
Chats	85
Cookie-Names	438
Cryptocurrencies	641
Cves	121
Decodeds	1902
DomHashes	8499
Domains	1580
Etags	5877
Favicons	778
File-Names	64
HHashes	689
Images	434
Ocrs	198
PGP Dumps	144
Qrcodes	111
User-Accounts	337
Usernames	498

Up Down Queue UP DOWN 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/ywyh44v7xoccwueflc6-8339-40a5-937e-e84b7ae351fb	infoleak:automatic-detection:"bitcoin-address"	16:26:01

TLP:CLEAR

20/31

Search

The screenshot shows a web-based search interface with two main sections: 'Tor and Web Search:' and 'Chats Search:'.

Tor and Web Search:

- Type:** Buttons for Tor, Web, and All. The 'Tor' button is selected.
- Content:** A search bar containing "Tor" and a dropdown menu showing "content to Search".
- Search Options:** A blue search button and an information icon.

Search Domain by name:

- Domain name:** An input field with a dropdown menu showing "content to Search".
- Filter:** A toggle switch between Onion Domains and Web Domains. The "Onion Domains" option is selected.
- Search Options:** A blue search button and an information icon.

Titles Search:

- Content:** A search bar containing "Content Search" and a dropdown menu showing "ID or content to Search".
- Options:** A toggle switch between Case Sensitive and Case Insensitive. The "Case Sensitive" option is selected.
- Search Options:** A blue search button and an information icon.

Chats Search:

- Type:** Buttons for discord, matrix, telegram, and All Chats. The 'discord' button is selected.
- Content:** A search bar containing "discord" and a dropdown menu showing "content to Search".
- Search Options:** A blue search button and an information icon.

Usernames Search:

TLP:CLEAR

21/31

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"

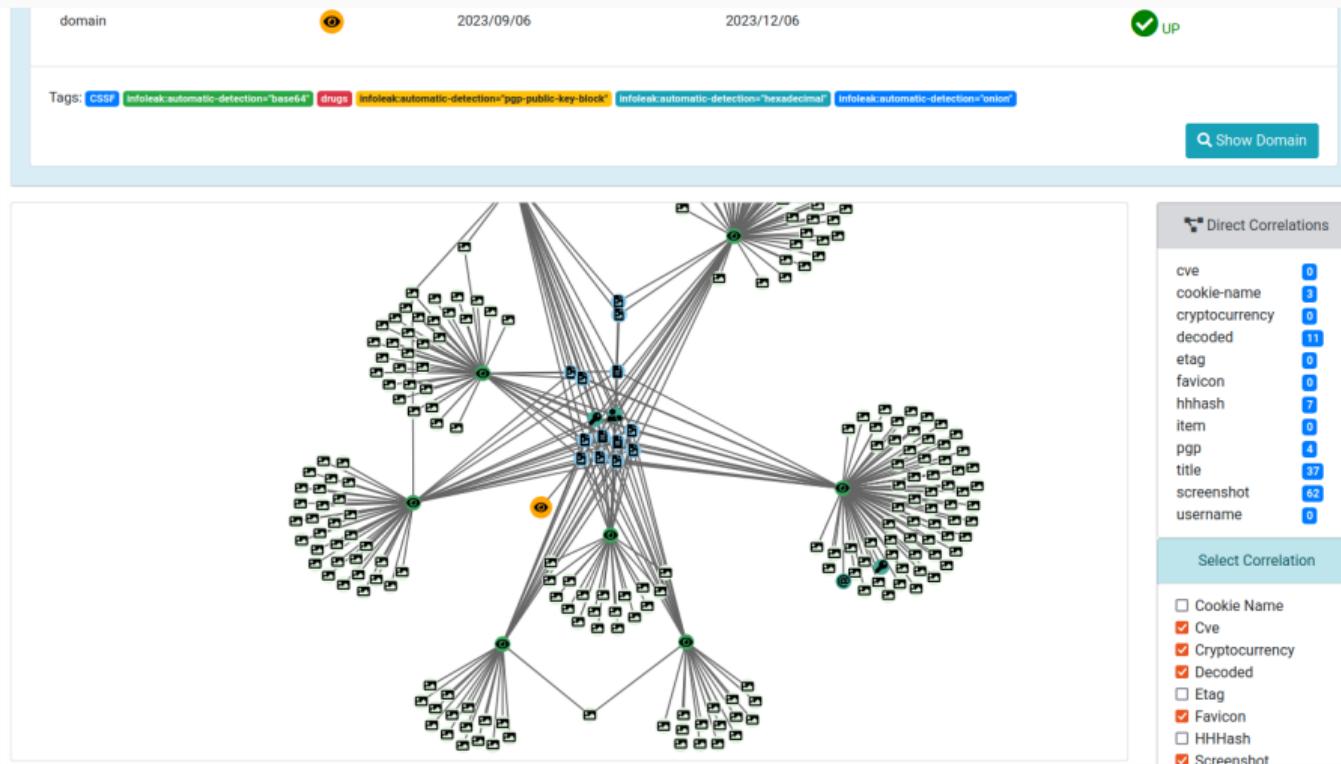
Showing 1 to 3 of 3 entries

Previous 1 Next

TLP:CLEAR

23/31

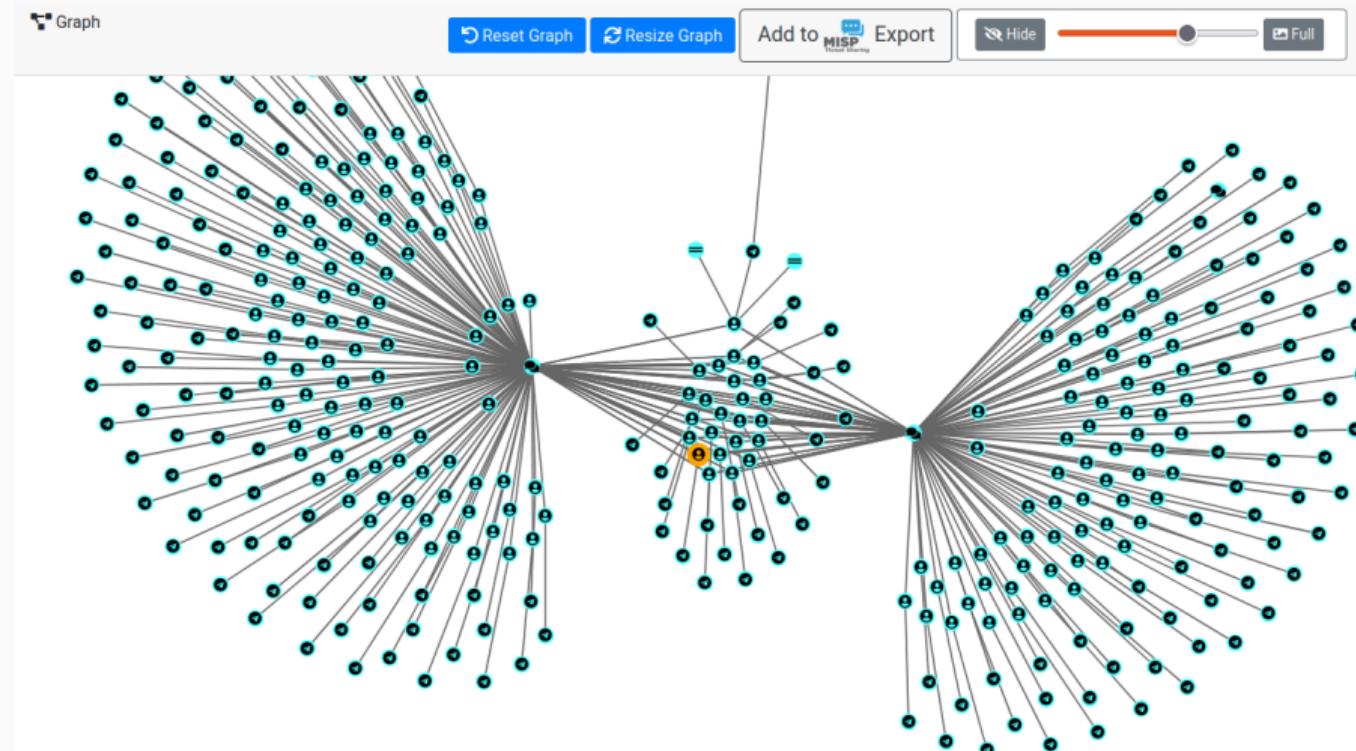
Correlations and relationship



TLP:CLEAR

24/31

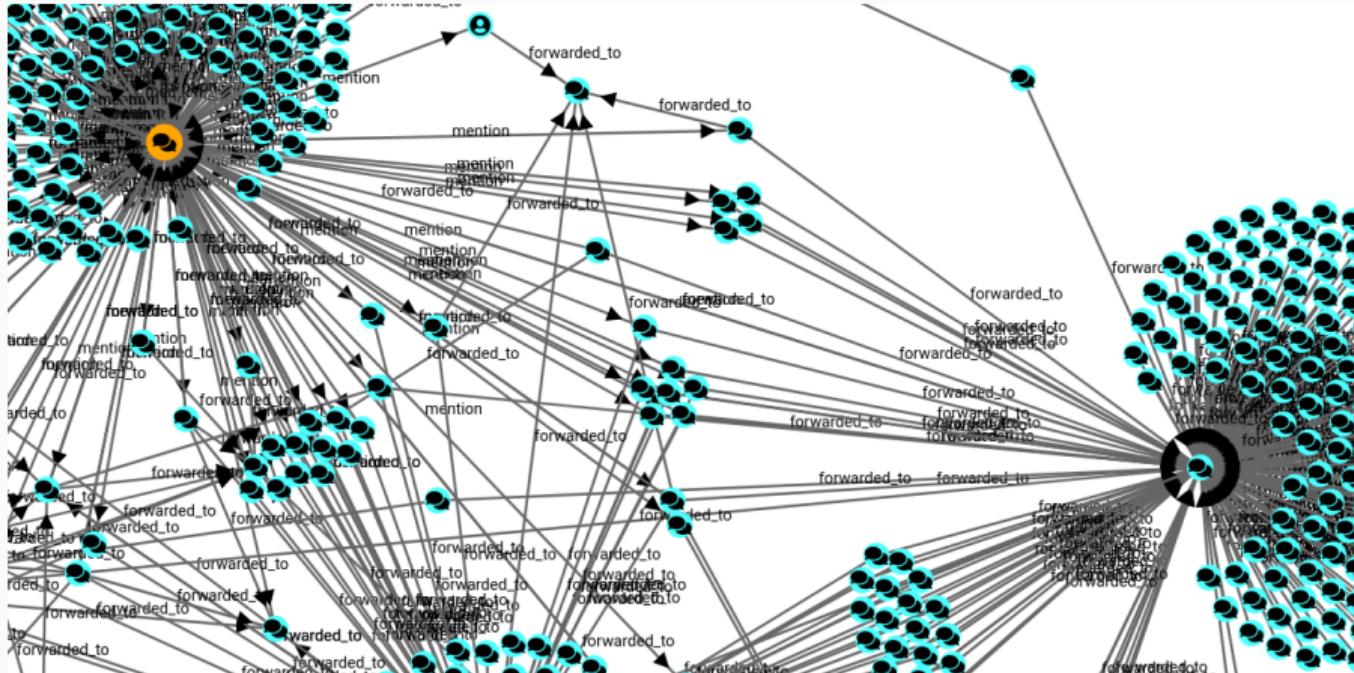
User Correlation - Common Chats



TLP:CLEAR

25/31

What are the Relationships Between Chats Group?



Investigations

Tor Coin Mixer	
UUID	9189d0e7cd4c47a29f85666e9507e0a5
Creator	admin@admin.test
Tags	<code>dark-web-topic="mixer"</code>
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	jamblerly2zpknhjbn3mhfdajmyddqbxuf1voa32h5w4o6ux3cqrd.onion	<code>infidak-automatic-detections="index"</code> <code>infidak-automatic-detections="pgp-public-key-block"</code>	
onion	bitmonhf4cpncluhwfifuskk23tvowsbwbe4tlthree74oxjmz2yyqqd.onion	<code>infidak-automatic-detections="index"</code>	
key	0x0B280956F0E7CAF		
mail	support@jambler.io		
telegram	jambler		
name	Jambler.io		

Showing 1 to 6 of 6 entries

TLP:CLEAR

Previous Next

27/31

- 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

- **Improved Mail Search – Next Release**
- **Lacus crawler improvements:** proxy selection, local cache, and cookie state transfer
- **Distributed Image Description Processing** - Automatic load balancing and per-domain classification across multiple Ollama servers
- **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 project⁷ : <https://github.com/ail-project/ail-framework>
- For questions, contact: info@circl.lu

⁷All 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.