

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

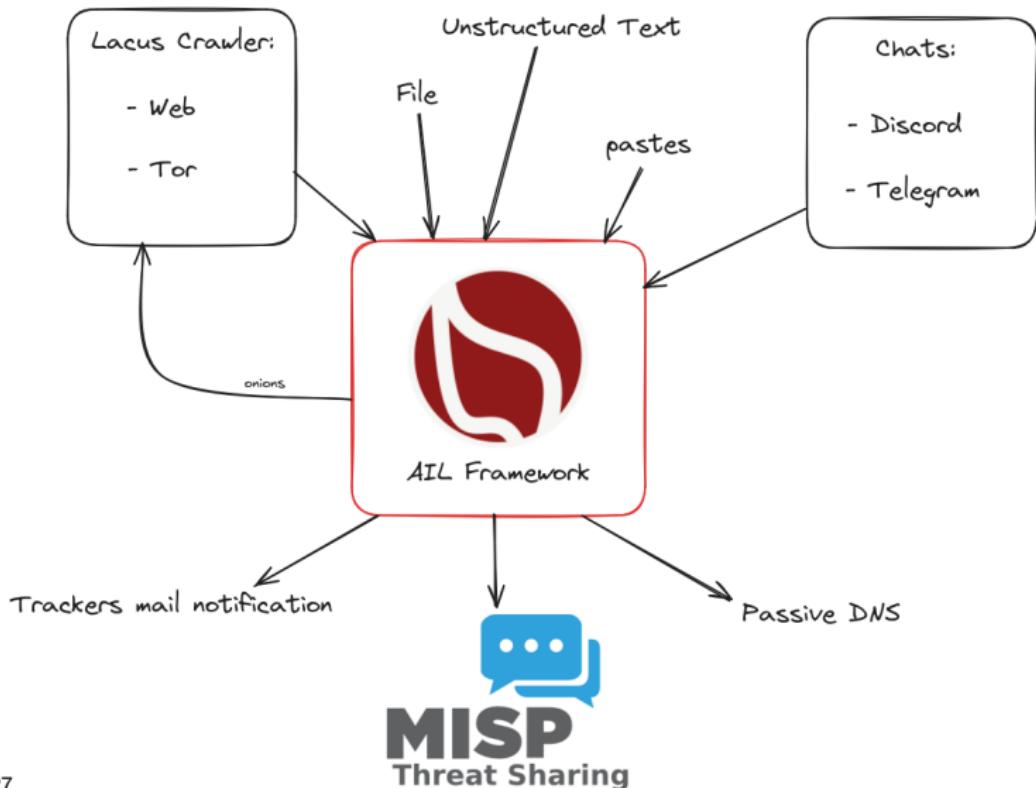
- Over the past years, CIRCL has developed the AIL project¹ to fulfill our needs at CIRCL in intelligence gathering and analysis.
- AIL features an extensible Python-based framework for the **analysis of unstructured information**, collected either through an advanced crawler manager or from various 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/>

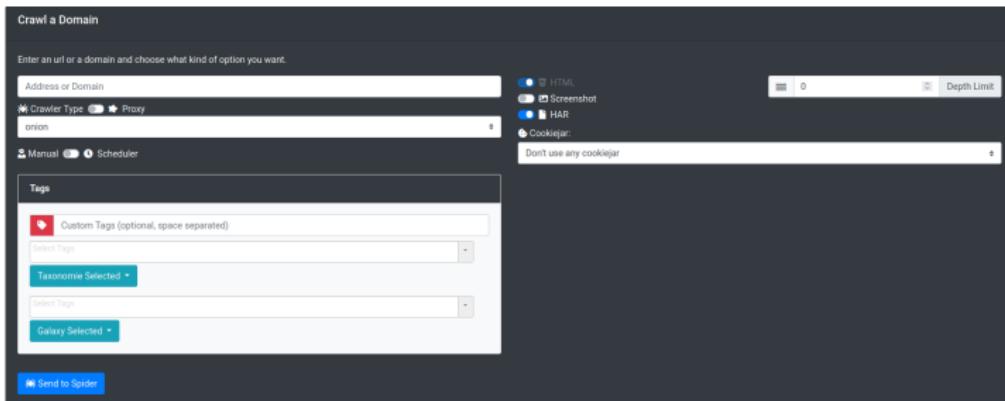
²<https://github.com/ail-project>

High Level Overview



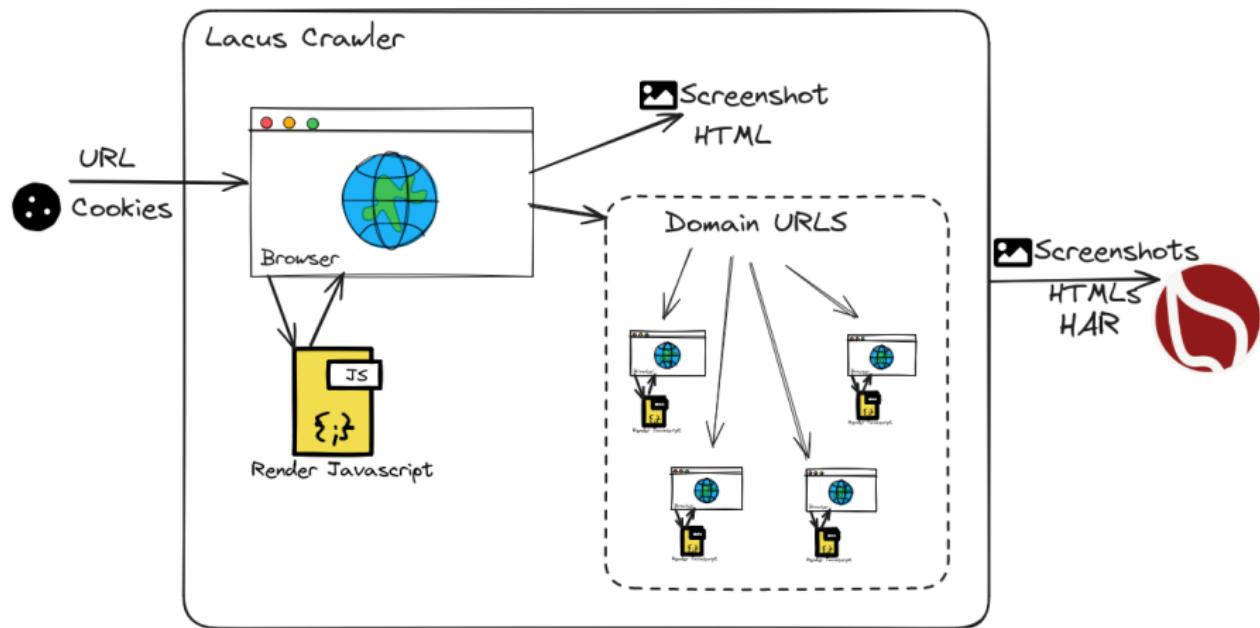
Collection - automate crawling

- 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 Delete

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": "047ab0cd97fff5bcc77edb6a"  
}
```

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

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

Crawler: Cookiejar

3thxemke2x7hcibu.onion : X 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 129

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, zulopri. 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 SECTION

Shere Khan - Official Forum
Private Messages

Inbox Compose Message Manage Folders Empty Folders Download Messages 1% of PM space used.

Message Title Sender Date/Time Sent [sec]

KEY FOR PRIVATE SECTION Jok3 3 hours ago

Verification Jok3 03-09-2020, 11:55 AM

Move To Inbox Delete the selected messages

Your Profile Edit Profile Change Password Change Email Change Avatar Change Signature Edit Options

Miscellaneous Group Memberships BuddyIgnore List Manage Attachments Saved Drafts Subscribed Threads Forum Subscriptions View Profile

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>

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: [8](#)

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

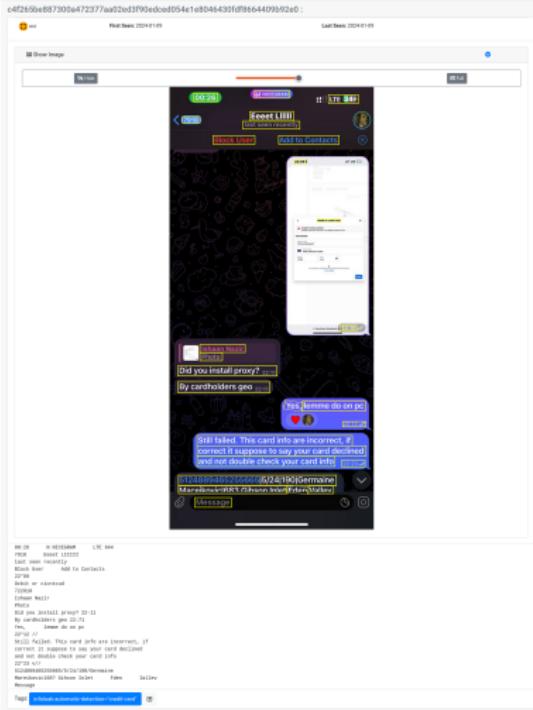
Sub-Channels:

Show [10](#) [+ entries](#)

Search:

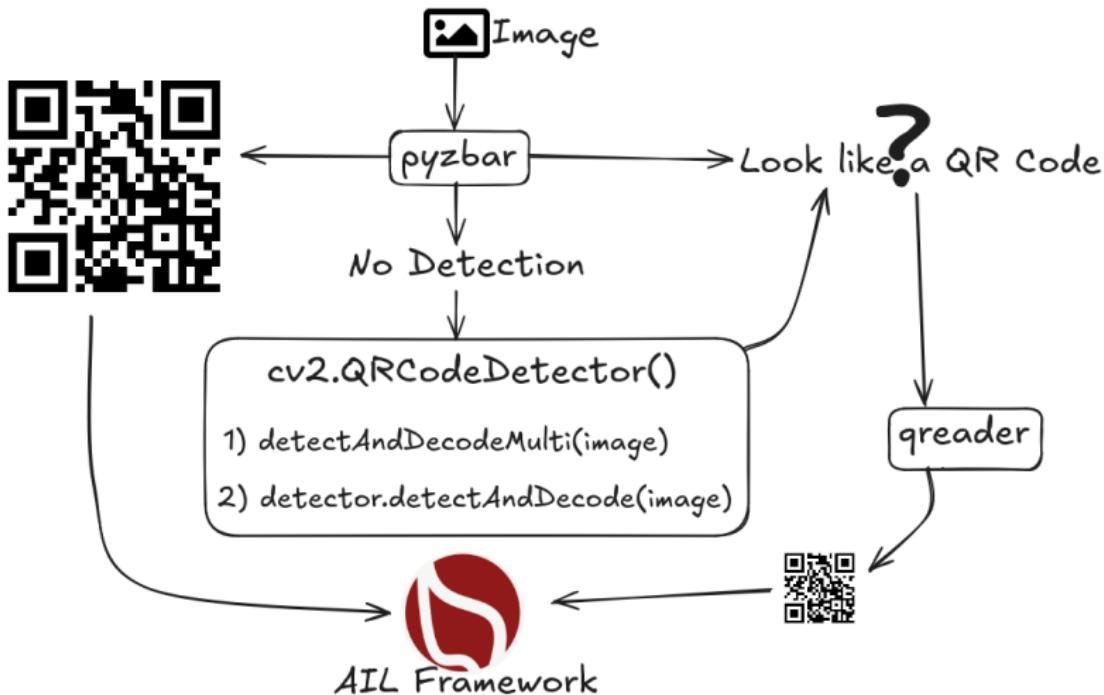
Icon	Name	ID	Created at	First Seen	Last Seen	...
	General	2125229770/1	2024-03-07 06:43:47	2024-03-07	2024-04-12	4121

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



QR

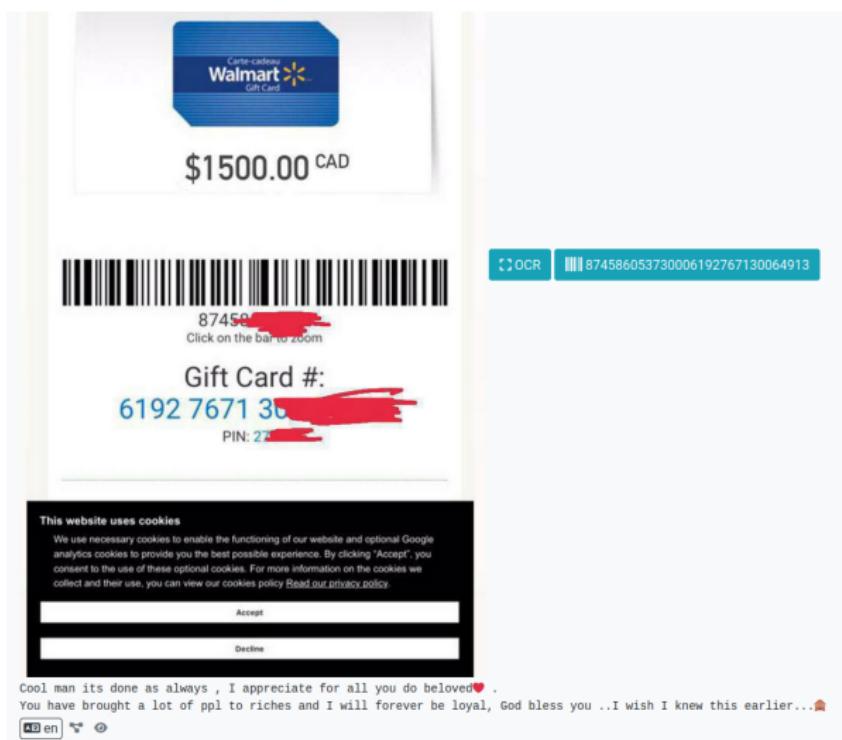
likeaboss



000201010211021643327300096494304155290090009649415312676076426760764000
Mai62360520AP9J3I00000000000000000708447583166304F69D

Someone with union pay link or QR code
Hit me up

Bar Codes



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 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	0
Chats	85
Cookie-Names	428
Cryptocurrencies	641
Cves	13
Decodeds	1902
DomHashes	8409
Domains	1580
Etags	5877
Favicons	778
File-Names	64
HHHashes	969
Qrcodes	111
Images	454
Ocrs	118
PGP Dumps	144
User-Accounts	337
Titles	5895
Usernames	498

Metrics

Up	1576
Down	3107
Crawled	4665
Queue	14916
UP	4
DOWN	9
Crawled	4
Queue	0

Logs

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

Leaks

ID	Tags	Time
crawled/2024/12/07/ywly44v7ixocqvuelfc6-h339-40a5-937e-e84b7ae351fb	infoleak submission/crawler	16:26:01

Page Footer

16 of 27

Search

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

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 8](#) [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
        $open_cert = "BEGIN CERTIFICATE" wide ascii nocase
        $pkcs7 = "BEGIN PKCS7"

    condition:
        any of them
}
```

2023-05-12 2023-05-31

[Tracked Objects](#)



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

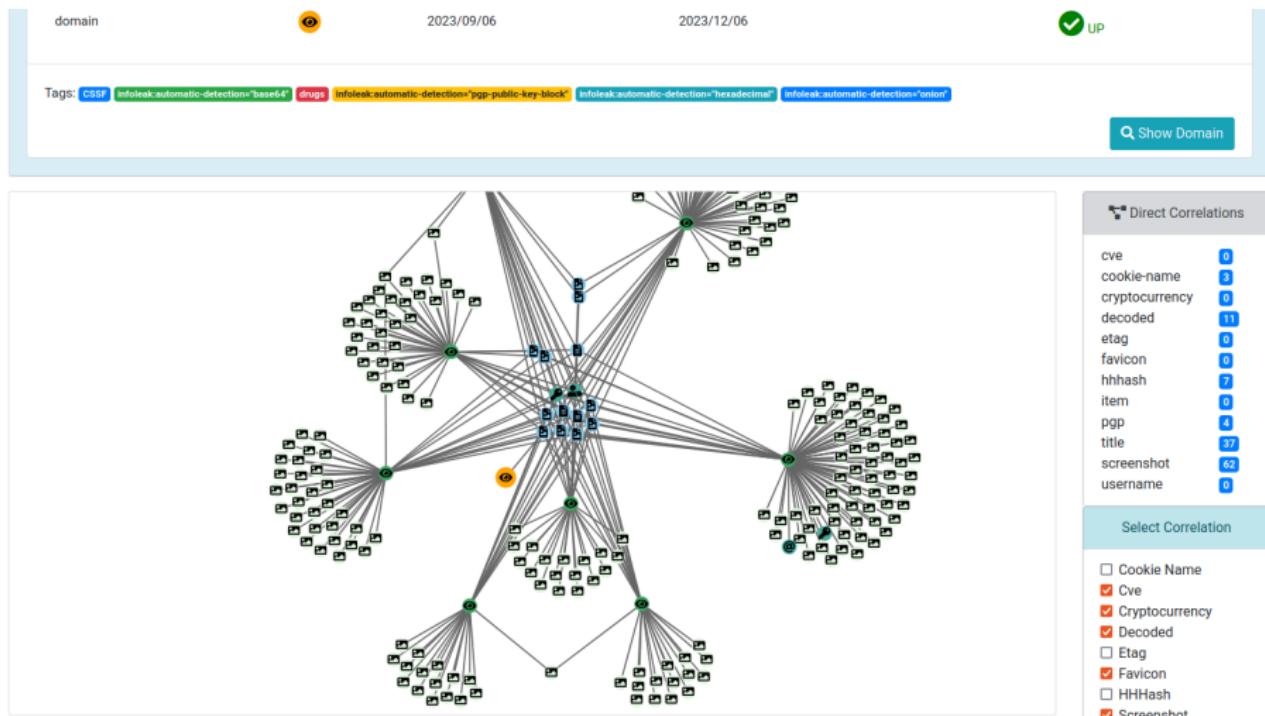
```
rule certificates
{
    meta:
        author = "@KevTheHermit"
        info = "Part of PasteHunter"
        reference = "https://github.com/kevthehermit/PasteHunter"

    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
        $open_cert = "BEGIN CERTIFICATE" wide ascii nocase
        $pkcs7 = "BEGIN PKCS7"

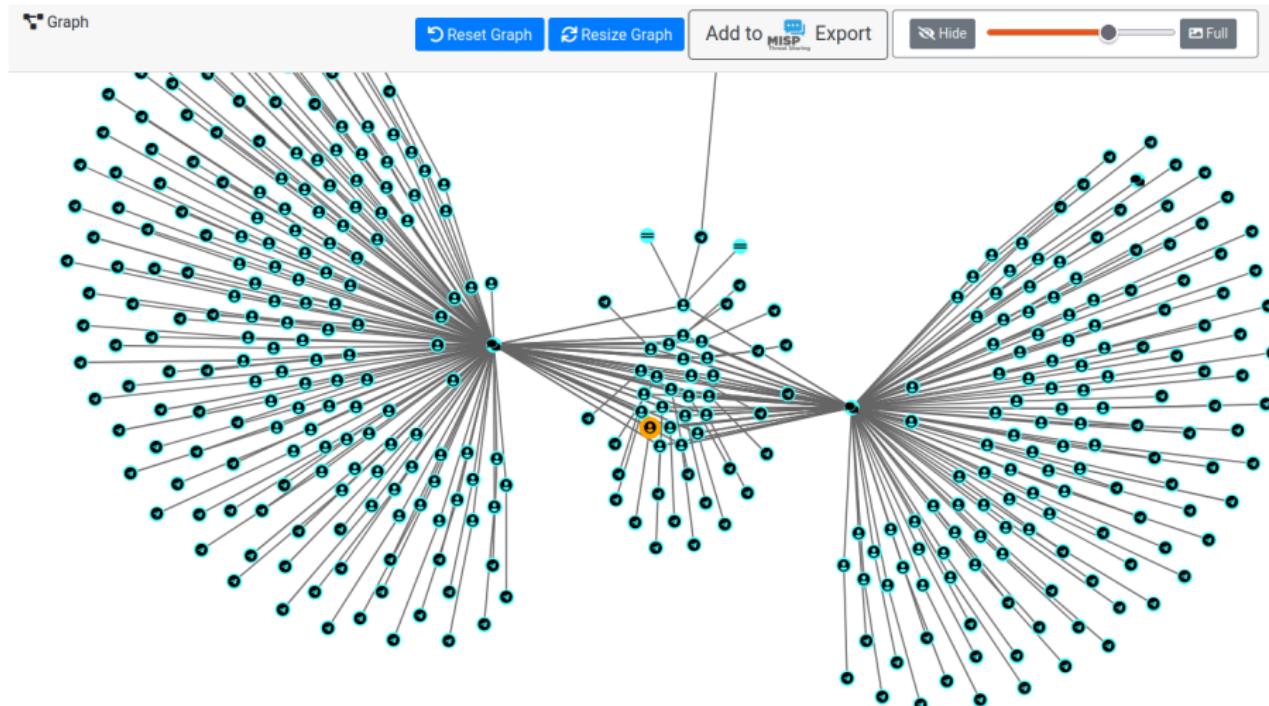
    condition:
        any of them
}
```

Type	T1	T2	Id	Tags	T3	T4
●			archive/gist.github.com/2023/04/14/laizmiranda7_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.automatic-detection="private-key" test12 infoleak.automatic-detection="rsa-private-key" infoleak.automatic-detection="pgp-static-key" test23 infoleak.automatic-detection="certificate" infoleak.automatic-detection="hexen"		
●			archive/gist.github.com/2023/04/13/chipzoller_d86d2d737d02ad4fe9d30a897170761.gz	test12 test23 infoleak.automatic-detection="certificate"		

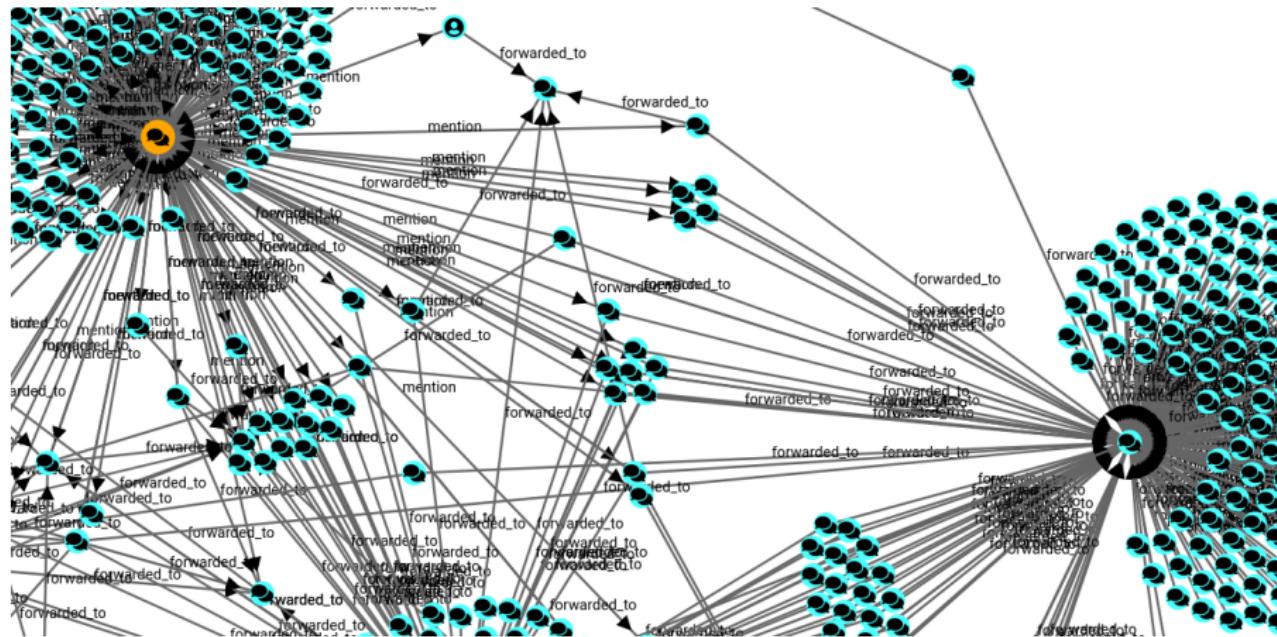
Correlations and relationship



User Correlation - Common Chats



What are the Relationships Between Chats Group?



Investigations

Tor Coin Mixer	
UUID	9189d0e7c04c47a29f85666e9507e0a5
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

[Delete](#) [Edit](#) [Export as Event](#)

Objects

Show 10

Search:

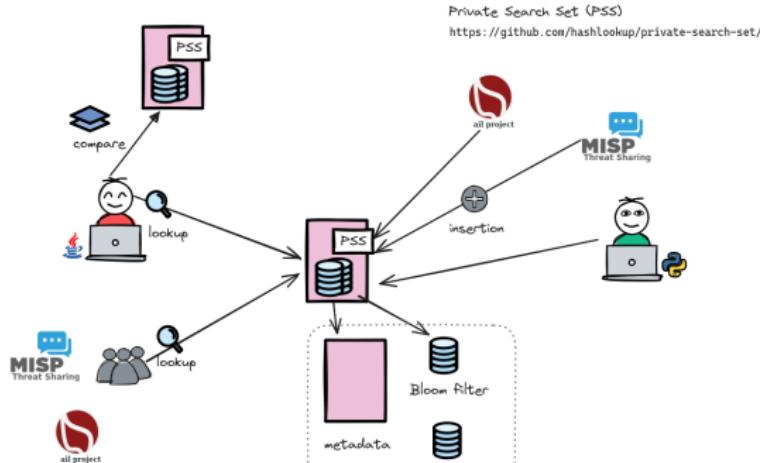
Type	Id	Tags	
onion	jambler7zgpxnhjlmj3mrfajymyddqxbufl6voa32h5w4o0x3crqd.onion	<code>info:ak-automatic-detection="true"</code> <code>info:ak-automatic-detection="pgp-public-key-block"</code>	🔗
onion	b7mohh74cpncluhwfllsk23tvowswbe4tlldree74oxjmz2yyqqd.onion	<code>info:ak-automatic-detection="true"</code>	🔗
key	0xD0B2B0956F0E7CAF		🔗
mail	support@jambler.io		🔗
telegram	jambler		🔗
name	Jambler.io		🔗

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
- MISP export with new correlation types
- Automatic geolocation
- Lacus improvement including proxy selection, local cache and cookies state transfer
- 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

