

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**.

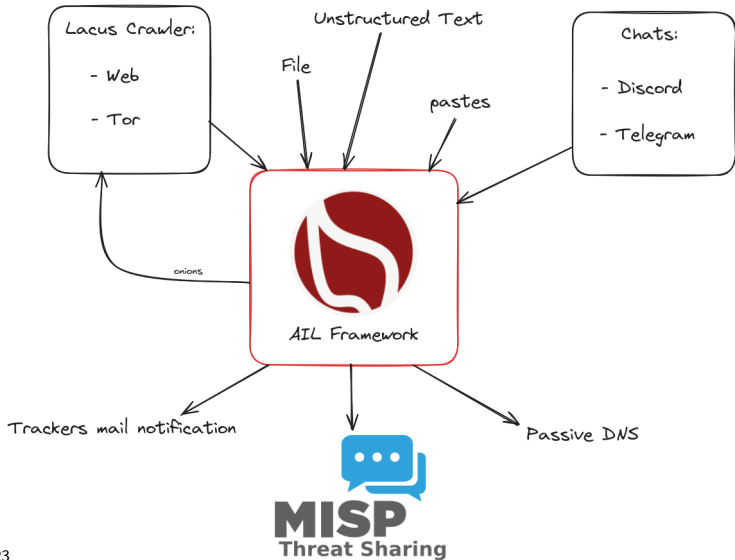


ail project

¹<https://www.ail-project.org/>

²<https://github.com/ail-project>

High Level Overview



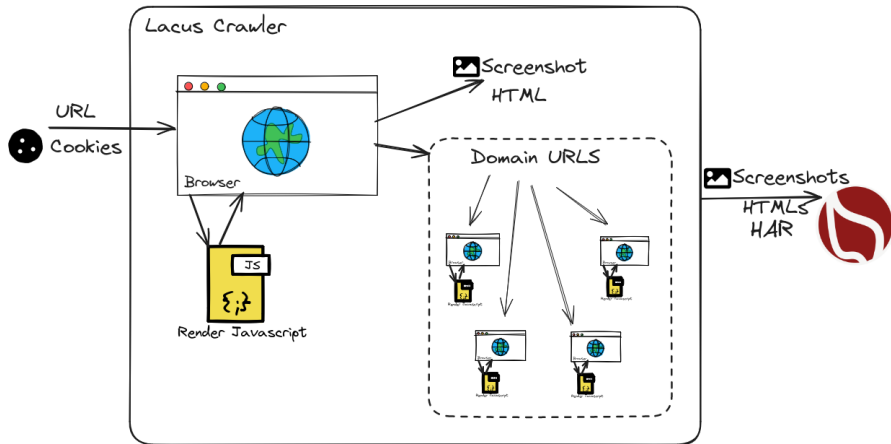
Collection - automate crawling

- 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**.

The screenshot shows the 'Crawl a Domain' interface. At the top, it says 'Enter an url or a domain and choose what kind of option you want.' Below this is a text input field for 'Address or Domain'. To the right of this field are three radio buttons: 'HTML' (selected), 'Screenshot', and 'HAR'. Further right is a 'Depth Limit' input field. Below the 'Address or Domain' field is a 'Crawler Type' section with a dropdown menu showing 'python'. To the right of this is a 'Cookiejar' section with a dropdown menu showing 'Don't use any cookiejar'. Below the 'Crawler Type' section is a 'Scheduler' section with two radio buttons: 'Manual' (selected) and 'Scheduler'. Below the 'Scheduler' section is a 'Tags' section with a text input field for 'Custom Tags (optional, space separated)' and two dropdown menus: 'Tazomine Selected' and 'Galaxy Selected'. At the bottom left of the interface is a blue button labeled 'Send to Spider'.

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.onioi	2020/03/31	90674deb-38fb-4eba-a661-18899ccb3841	admin@admin.test

Edit Description

Add Cookies

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

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

```
{
  "domain": ".3thxemke2x7hcibu.onioi",
  "name": "sid",
  "path": "/forum/",
  "value": "847ab8cd97ff5bcc77eddb6a"
}
```

```
{
  "name": "remember_token",
  "value": "12158cddd151d74d341f23"
}
```

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

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 Export

Decoded 1

Screenshot 134

Crawled Items

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

Show 10 entries

Search:

Crawled Pastes

Hide

Full resolution

Shere Khan

Portal Search Member List Help

Welcome back, zuluport. 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 Jack3 (ID: KEY FOR PRIVATE SECTIONS)

Shere Khan - Official Forum

Private Messages

Home

User CP Home

Messages

Compose

Inbox

Send

Send

Trash Can

Tracking

Build Folder

Your Profile

Get Profile

Change Password

Change Email

Change Avatar

Change Signature

Build Options

MicroNamex

Group Memberships

Buddy/Ignore List

Manage Attachments

Saved Drafts

Subscribed Threads

Forum Subscriptions

View Profile

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

Enter Keywords Search PMs (Advanced Search)

Message Title Sender Date/Time Sent (asc)

KEY FOR PRIVATE SECTIONS Jack3 3 hours ago

Verification Jack3 03-09-2020, 11:55 AM

Move To Inbox or Delete the selected messages

Jump to Folder: Inbox Get

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

Powered by MyBB, © 2002-2020 MyBB Group.


Current time: 03-23-2020, 01:33 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:

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

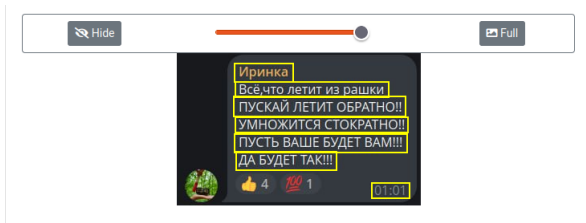
Sub-Channels:

Show 10 s 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

Processing - OCR



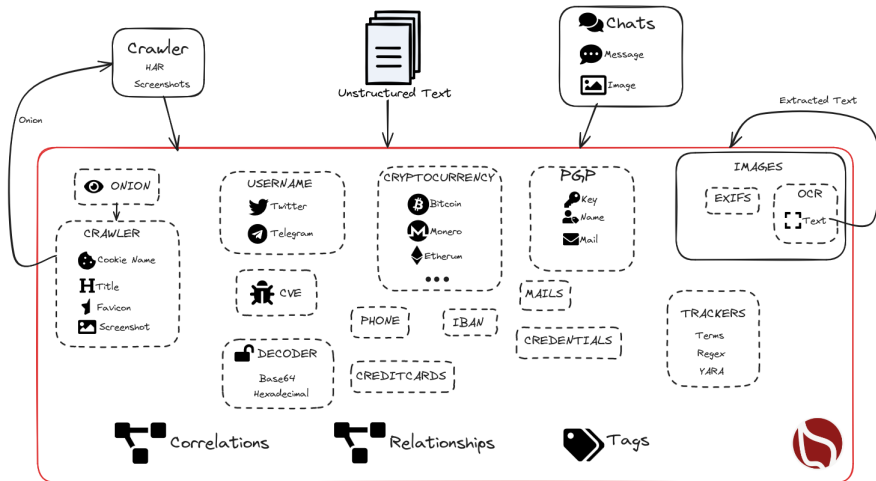
Иринка
Всё, что летит из рашки
ПУСКАЙ ЛЕТИТ ОБРАТНО!!
УМНОЖИТСЯ СТОКРАТНО!!
ПУСТЬ ВАШЕ БУДЕТ ВАМ!!!
ДА БУДЕТ ТАКИ
01.01

Irinka
Everything that flies from Russia
LET IT FLY BACK
IT WILL MULTIPLY A HUNDRED TIMES!
LET YOURS BE YOURS!
LET IT BE SO
01.01

ru

Tags: +

Analysis of unstructured information



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>

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

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 0
item 00

[Edit Tracker](#)  

Yara Rule:

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

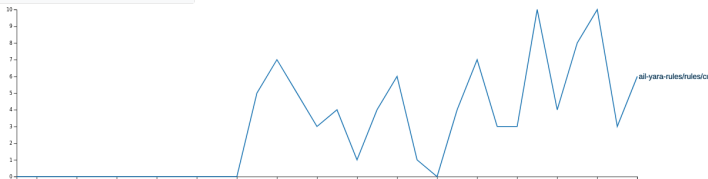
  strings:
    $ssh_priv = "BEGIN RSA PRIVATE KEY" wide ascii nocase
    $openssh_priv = "BEGIN OPENSSH PRIVATE KEY" wide ascii nocase
    $dsa_priv = "BEGIN DSA PRIVATE KEY" wide ascii nocase
    $ec_priv = "BEGIN EC PRIVATE KEY" wide ascii nocase
    $pgp_priv = "BEGIN PGP PRIVATE KEY" wide ascii nocase
    $pem_cert = "BEGIN CERTIFICATE" wide ascii nocase
    $pkcs7 = "BEGIN PKCS7"

  condition:
    any of them
}
```

 2023-05-12

 2023-05-31

 [Tracked Objects](#)



Retro Hunt

test completed

Date2023/05/10

DescriptionNone

Tags

Creatoradmin@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"

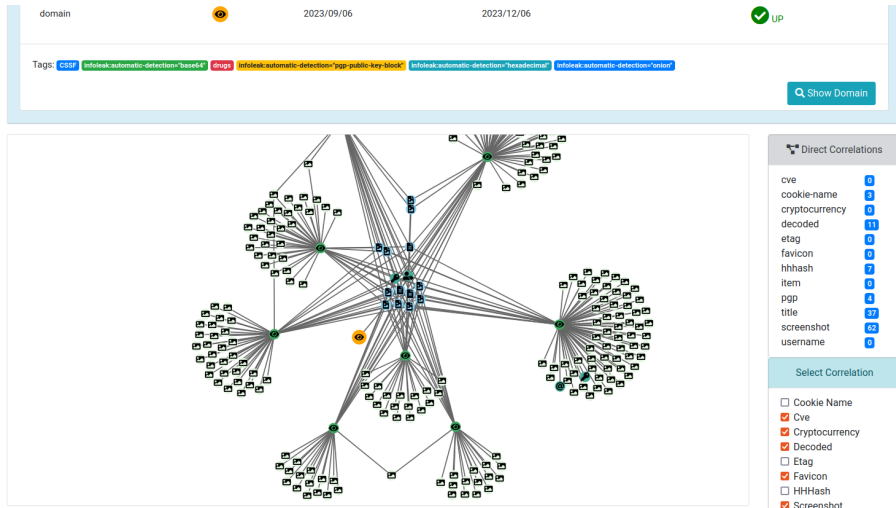
  condition:
    any of them
}
```

Show 10 entries

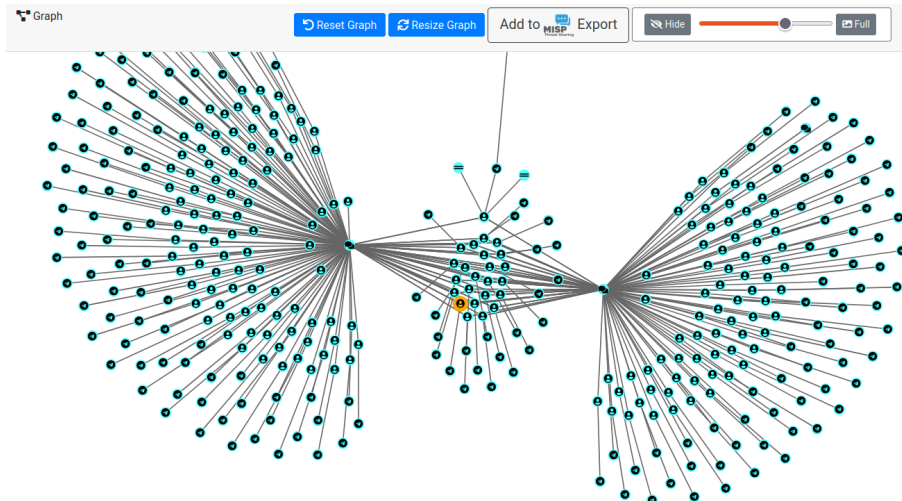
Search:

Type	Id	Tags
	archive/gist.github.com/2023/04/14/huizmiranda7_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"

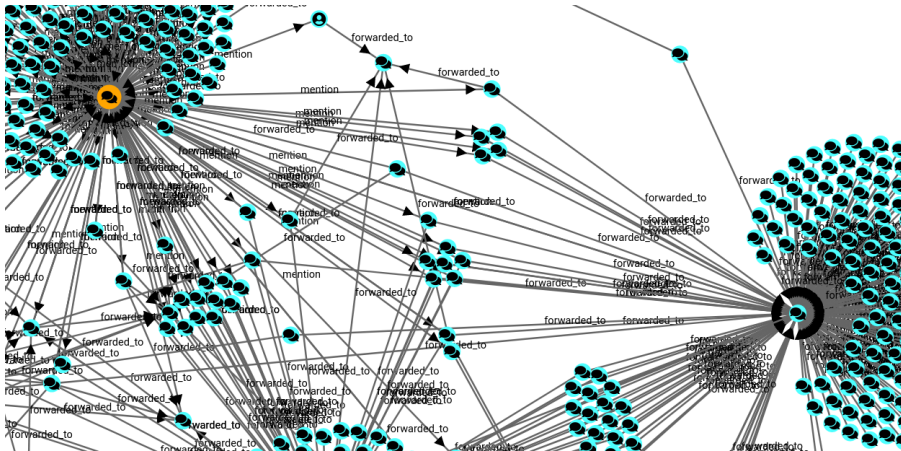
Correlations and relationship



User Correlation - Common Chats



What are the Relationships Between Chats Group?



Investigations

Tor Coin Mixer

UUID	9189d0e7c04c47a2985666e9507e0a5
Creator	admin@admin.test
Tags	dark-web:torpcr: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

Delete

Edit

Export as Event

Objects

Show 10 entries

Search:

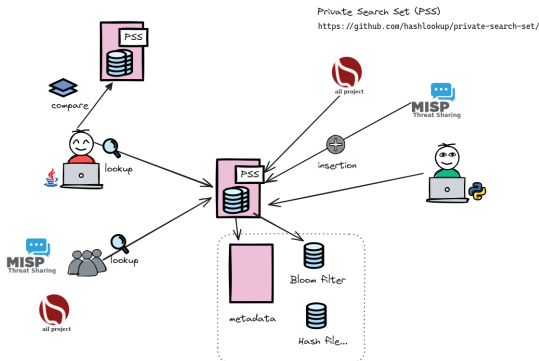
Type	Id	Tags	
onion	jambler72gpknhmj3mhdaajmyddqxbuif6vos32h5w4otux3crqd.onion	infoleak-automatic-detection==anon infoleak-automatic-detection==ygg-public-key-block	<div></div>
onion	btmxihtf4cpcncluhwflussk23ltvowswebe4tridree74qxjnz2vyqgd.onion	infoleak-automatic-detection==anon	<div></div>
key	0xD3B280956F0E7CAF		<div></div>
mail	support@jambler.io		<div></div>
telegram	jambler		<div></div>
name	Jambler.io		<div></div>

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**
- Improved filtering in dashboard
- **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**

