

# AIL Project

How to Improve and Support Your Threat Intelligence Process



**CIRCL**

Computer Incident  
Response Center  
Luxembourg

Alexandre Dulaunoy

[alexandre.dulaunoy@circl.lu](mailto:alexandre.dulaunoy@circl.lu)

[info@circl.lu](mailto:info@circl.lu)

November 8, 2023

# Background

---

- Over the past five years, we have developed the AIL project<sup>1</sup> to fulfill our needs at CIRCL in intelligence gathering and analysis.
- As AIL gained popularity, an increasing number of users began integrating it into their **threat intelligence processes and workflows**.
- In this presentation, we outline some of the processes where AIL can serve as a valuable tool, **facilitating and enhancing the work of intelligence analysts**.

---

<sup>1</sup><https://www.ail-project.org/>

## AIL overview

---

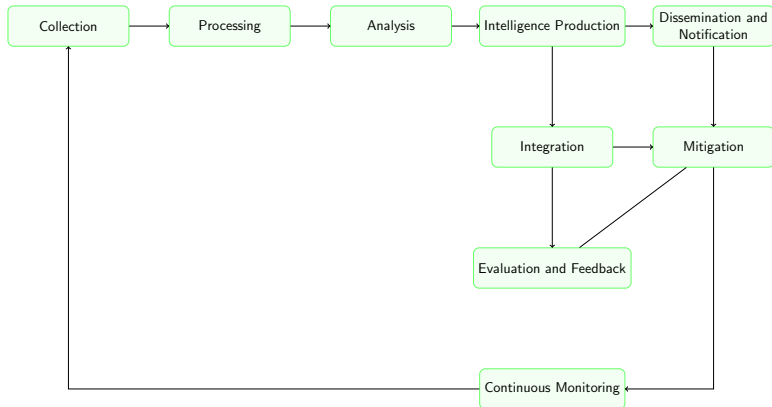
- The AIL Project is an open-source framework<sup>2</sup> comprising various modules designed for the **collection, crawling, digging, and analysis of unstructured data**.
- AIL features an extensible Python-based framework for the **analysis of unstructured information**, collected either through an advanced Crawler manager or from various feeders, including social networks and custom feeders.
- AIL also provides support for actively **crawling Tor** hidden services, as well as crawling protected websites and forums by utilizing pre-recorded session cookies.

---

<sup>2</sup><https://github.com/ail-project>

# Threat Intelligence Process at CIRCL

---



## Common questions from constituents

---

- Do you **know if we are a target** of this adversary group?
- We have **observed a partnering company experiencing a ransomware incident**, and we are concerned about the impact on our organization.
- Can you determine if our **sector is a target** of this threat actor?
- Have you come across phishing kits targeting our bank/service or any instances of our **data being stolen** on the "dark web"?

## Challenges and opportunity

---

- **Reducing repetitive tasks** for the analysts.
- **Preparing factual intelligence evidence** for intelligence production, including human-readable reports and MISP structured intelligence.
- **Correlating information** from multiple sources, especially when different analysts are working with different sources on their end.
- **Facilitating the integration** of "intelligence requests" from our constituents.

# 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, etc.) that can be used to subscribe to chat channels.
- All the **collected messages are then processed and analyzed** within the AIL's *processing* and *analysis* stages.

DDosia Project :





1228309110

Icon	Name	ID	First Seen	Last Seen	NB Sub-Channels
	DDosia Project	1228309110	2023-10-20	2023-11-06	5

## Sub-Channels:

Show 10 entries

Search:

Icon	Name	ID	First Seen	Last Seen	NB Messages
	Общий чат	<a href="#">1228309110/1</a>	2023-10-20	2023-11-06	1498
	Полезные материалы	<a href="#">1228309110/34221</a>	2023-10-21	2023-11-06	360
	DDosia - поддержка	<a href="#">1228309110/34219</a>	2023-10-20	2023-11-05	417
	Предложение целей	<a href="#">1228309110/34217</a>	2023-10-24	2023-11-05	26

## Collection - automate crawling

---

- Crawling can be a challenging task, for example, gathering all the blog posts from ransomware groups<sup>3</sup>, which can be demanding for an analyst.
- ALL offers a crawling feature that can **initiate regular crawls using a standard spawned browser.**

The screenshot shows the 'Crawl a Domain' configuration window. It includes a text input for 'Address or Domain', a 'Crawler Type' dropdown set to 'onion', and a 'Scheduler' section with 'Manual' selected. A 'Tags' section contains a 'Custom Tags' input and two dropdown menus with 'Taxonomie Selected' and 'Galaxy Selected'. On the right, there are radio buttons for 'HTML', 'Screenshot', 'HAR', and 'CookieJar', with 'Screenshot' selected. Below these is a 'Depth Limit' input and a 'CookieJar' dropdown set to 'Don't use any cookiejar'. A 'Send to Spider' button is at the bottom left.

---

<sup>3</sup><https://www.ransomlook.io/>





# Processing - deduplicating source/information

- When collecting data from numerous sources, encountering duplicate information is common, and distinguishing between them can be challenging.
- AIL's correlation between page titles, screenshots, and HTTP headers matching helps **identify copy-cat sources**.

vm474ll3aqnelbyxwlljp2opswzeow7dkl6xwkc4hej6oeqcyshzdad.onion :

Object type	type	First seen	Last check	Port	Status
domain		2023/11/08	2023/11/08		UP

Tags:

[Show Domain](#)

[Investigations](#)

Graph [Reset Graph](#) [Shrink Graph](#) Add to [Map](#) Export [Full](#)

Graph Incomplete, Max Nodes Reached

Select Correlation

- Cookie Name
- Cve
- Cryptocurrency
- Decoded
- Etag
- HttHash
- Screenshot
- Title
- PGP
- Username
- Domain
- Item

Correlation Depth:

Max number of nodes:

# Analysis - automatic detection from collection

- Processing automatically collected information can be a challenging task.
- ALL processes all the collected items for any **hunting rules** and **utilizes MISP taxonomies to tag the matching information.**

```
vargakrisztian44@gmail.com:9320111 |Plan = GPT Plus  
vebzbao@outlook.com:Sugar21212 |Plan = GPT Plus  
vivek@eyuva.com:$EyuvaSubhas009 |Plan = GPT Plus  
wahomeematah13@gmail.com:wahomee100 |Plan = GPT Plus  
walterbeyn@hotmail.com:Tutubye1-19 |Plan = GPT Plus  
waqasahadwik192727@gmail.com:incorrect192 |Plan = GPT Plus  
wardnj0720@gmail.com:Mickey52 |Plan = GPT Plus  
web.acowebexperts@gmail.com:Mac08seok |Plan = GPT Plus  
yanaili@yahoo.com:Panchatra2022... |Plan = GPT Plus  
yano@chagchef.com:Kyeft914717-1991 |Plan = GPT Plus  
yellowtreecanada@gmail.com:Mustanggt120195 |Plan = GPT Plus  
yulihaoo2007@gmail.com:Euro15421 |Plan = GPT Plus  
z.s.marcos.j@gmail.com:ToBeMlllionare |Plan = GPT Plus  
zahrarirvi813@gmail.com:sana1jaz05 |Plan = GPT Plus  
Zaibaa.pathan@gmail.com:Dandelion1231 |Plan = GPT Plus  
zainlynx06@gmail.com:buratako123 |Plan = GPT Plus  
zellloppj@gmail.com:Chocolati1 |Plan = GPT Plus  
zordope@gmail.com:Lisa2019 |Plan = GPT Plus
```

Don't change password or email.....

**Infomax** automatic-detection+root **Infomax** automatic-detection+credential  
" " @



1427620096  
Camera  
5472920900330xxx

# Analysis - evaluating vulnerability severity/risk

- What is the visibility, usage, mentions, or risk of a vulnerability observed in forums, channels, pastes, or websites?
- AIL can assist you in determining the severity/risk level or in **reviewing the usage of a vulnerability** (e.g., the number of PoCs).

The screenshot displays the details for CVE-2023-46604. At the top, a table lists the object type as 'cve', with a 'First seen' date of 20231104, a 'Last seen' date of 20231104, and 'NB seen' as 2. Below this, a summary states: 'Apache ActiveMQ is vulnerable to Remote Code Execution. The vulnerability may allow a remote attacker with network access to a broker to run arbitrary shell commands by manipulating serialized class types in the OpenWire protocol to cause the broker to instantiate any class on the classpath. Users are recommended to upgrade to version 5.18.10, 5.16.7, 5.17.6, or 5.18.3, which fixes this issue.' Metadata includes 'Published' (2023-10-27T16:13:00), 'Modified' (2023-10-29T01:44:00), and 'last modified' (2023-10-29T01:44:00). A 'References' section is visible below. The bottom portion of the image shows a network graph interface with a 'Graph' title, a warning 'Graph Incomplete, Max Nodes Reached', and buttons for 'Reset Graph', 'Refresh Graph', 'Add to watch', 'Export', and 'Full'. The graph consists of numerous nodes, some red and some blue, connected by lines. A 'Select Correlation' sidebar on the right lists various attributes like 'Cookie Name', 'Cve', 'Cryptocurrency', 'Decoded', 'Etag', 'HHHash', 'Screenshot', 'Title', 'PGP', 'Username', 'Domain', and 'Item'. The 'Correlation Depth' is set to 2.

Object type	First seen	Last seen	NB seen
cve	20231104	20231104	2

**Summary** Apache ActiveMQ is vulnerable to Remote Code Execution. The vulnerability may allow a remote attacker with network access to a broker to run arbitrary shell commands by manipulating serialized class types in the OpenWire protocol to cause the broker to instantiate any class on the classpath. Users are recommended to upgrade to version 5.18.10, 5.16.7, 5.17.6, or 5.18.3, which fixes this issue.

**Published** 2023-10-27T16:13:00  
**Modified** 2023-10-29T01:44:00  
**last modified** 2023-10-29T01:44:00

**References**

**Tags**

**Investigations**

**Graph**  
Graph Incomplete, Max Nodes Reached

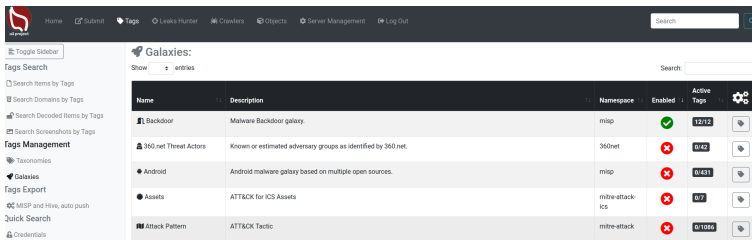
**Select Correlation**

- Cookie Name
- Cve
- Cryptocurrency
- Decoded
- Etag
- HHHash
- Screenshot
- Title
- PGP
- Username
- Domain
- Item

**Correlation Depth:** 2

# Analysis - Standardising labels and taxonomies

- Attribution and classification can be challenging for analysts. Facilitating integration with other tools, processes, and teams.
- **ALL leverages the entire MISP galaxy, including threat actor data, taxonomies, and the ability to assign tags to every item.**

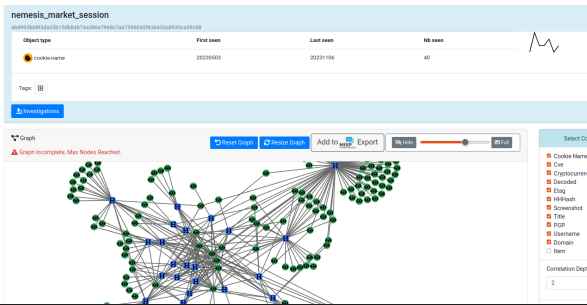


The screenshot displays the MISP Galaxies management interface. The top navigation bar includes links for Home, Submit, Tags, Leaks Hunter, Crawlers, Objects, Server Management, and Log Out. A search bar is located in the top right corner. On the left side, there is a sidebar with a 'Toggle Sidebar' button and a 'Tags Search' section containing various search filters like 'Search Items by Tags', 'Search Domains by Tags', 'Search Decoded Items by Tags', and 'Search Screenshots by Tags'. Below this is a 'Tags Management' section with options for 'Taxonomies', 'Galaxies', 'Tags Export', 'MISP and Hive, auto push', 'Quick Search', and 'Credentials'. The main content area is titled 'Galaxies:' and shows a table of galaxies. The table has columns for Name, Description, Namespace, Enabled, and Active Tags. The 'Enabled' column uses a green checkmark for 'Yes' and a red 'X' for 'No'. The 'Active Tags' column shows a count of active tags and a gear icon for configuration.

Name	Description	Namespace	Enabled	Active Tags
Backdoor	Malware Backdoor galaxy.	misp	Yes	12/12
360.net Threat Actors	Known or estimated adversary groups as identified by 360.net.	360net	No	0/42
Android	Android malware galaxy based on multiple open sources.	misp	No	0/431
Assets	ATT&CK for ICS Assets	mitre-attack-ics	No	0/7
Attack Pattern	ATT&CK Tactic	mitre-attack	No	0/1086

## Dissemination - distributing analysis

- AIL exports data using the **MISP standard format** and offers complete integration with MISP to facilitate the dissemination of data.
- All the context within AIL uses the **MISP taxonomies and galaxy**.
- The insights provided by AIL are often used as complementary information for threat intelligence reports and landscapes.



# Evaluation/Integration - review search rules on real dataset

- Reviewing matching rules on a large dataset, such as extensive ransomware leaks, can be cumbersome.
- AIL provides a "retro-hunt" functionality to search and **evaluate your YARA rules**.

The screenshot displays the AIL interface with two rule detail panels. The top panel shows the 'CSSF entities name search (mammouth)' rule, which is a YARA rule of type 'yara', tracked under 'custom-rules/2e3cbbb8-d093-49f4-9459-17138e38cae2.yar', dated 2023/09/12, with a global level. The bottom panel shows the 'rule nato' rule, also a YARA rule, with a meta description 'Part of all-yara-rules' and a reference to a GitHub repository. It includes a list of strings for detecting NATO-related terms and a condition that triggers on any of these strings.

**CSSF entities name search (mammouth)**

Type **yara**

Tracked custom-rules/2e3cbbb8-d093-49f4-9459-17138e38cae2.yar

Date 2023/09/12

Level Global

**Yara Rule:**

```
rule mammouth_CSSF {  
  
  meta:  
    description = "detect CSSF entities names"  
    author = "gallypette"  
  
  strings:  
    $x1 = "ALTRIO LINK" nocase  
    $x2 = "ARVIRA S.à r.l." nocase  
    $x3 = "Alsages S.à r.l." nocase  
    $x4 = "DB2 CONSULT" nocase  
    $x5 = "DOORNBOS LAGERVELD & PARTNERS S.À R.L." nocase  
    $x6 = "EM Wealth Office S.A." nocase  
    $x7 = "FC Consult S.A." nocase  
    $x8 = "FINANCE ET PATRIMOINE S.A." nocase
```

**NATO**

Date 2023/06/06

Description None

Tags [View tags](#)

Creator [aduba@cir.lu](#)

Filters {  
 "time": []  
}

Objects Match [View](#)

**rule nato**

```
{  
  meta:  
    author = "Bublar"  
    info = "Part of all-yara-rules"  
    reference = "https://github.com/all-graject/all-yara-rules"  
  
  strings:  
    $S1 = "COSMIC TOP SECRET" fullword wide ascii nocase  
    $S2 = "NATO SECRET" fullword wide ascii nocase  
    $S3 = "ATORNAL" fullword wide ascii nocase  
    $S4 = "NATO CONFIDENTIAL" fullword wide ascii nocase  
    $S5 = "NATO RESTRICTED" fullword wide ascii nocase  
    $S6 = "THIS DOCUMENT CONTAINS NATO CLASSIFIED INFORMATION" fullword wide ascii nocase  
    $S7 = "NATO UNCLASSIFIED - INTERNAL" fullword wide ascii nocase  
    $S8 = "Secret OTAN" fullword wide ascii nocase  
    $S9 = "Confidential OTAN" fullword wide ascii nocase  
    $S10 = "OTAN" fullword wide ascii nocase  
    $S11 = "Diffusion restreinte OTAN" fullword wide ascii nocase  
  
  condition:  
    1 of ($S*)
```

# Production collecting evidences

- Analysts need to gather evidence, insights, and intelligence to produce intelligence reports.
- AIL can support the creation of reports by offering a straightforward method to **organize discoveries for investigation.**

Father: [crawled/2023/10/01/zerodayhukmtc56zualcmtvtto5xfz7gytgt7poxgkmgegnq34p3xycyd.onion139a3549-5892-4dc1-91b3-4016e9b2c931](http://crawled/2023/10/01/zerodayhukmtc56zualcmtvtto5xfz7gytgt7poxgkmgegnq34p3xycyd.onion139a3549-5892-4dc1-91b3-4016e9b2c931)

The screenshot displays a web interface for a crawler. At the top, there are three buttons: "Add to MISP Export", "Investigations", and "Correlations Graph". Below these, the "Crawler" section shows the "Last Origin" as a list of URLs, including the one mentioned in the text above. A "Full resolution" button is visible next to the URL list. The main content area shows a preview of a webpage with a dark theme, featuring a hooded figure, a world map, and text including "Cracking", "0 Day", "Hack", "malware", "ransomware", and "Hack Learn".



## Improving internal capabilities

---

Whilst buying ready made intelligence is easy, you see here that going from a black box solution of questionable quality to something that you can vet and validate can be easily implemented - the costs will also be **invested in your internal experts rather than an opaque supplier.**

## Conclusion

---

- While AIL can be a valuable tool for **organisations dealing with data leaks and information breaches**, it's important to remember that it is primarily designed for information leak analysis and not for the entire threat intelligence process.
- Organizations should use **AIL in conjunction with other threat intelligence solutions** and processes to establish a comprehensive threat intelligence strategy.
- AIL is an open-source project, and if you discover modules that could assist in your processes, please let us know or contribute directly.
- Establishing **consistent and reproducible intelligence processes** throughout your organization.

## Links

---

- AIL project <https://github.com/ail-project> (**all components including feeders and crawler infrastructure**).
- AIL framework <https://github.com/ail-project/ail-framework> (**analysis framework**).
- Training materials and slide deck <https://github.com/ail-project/ail-training>.
- Co-funded by European Union under joint threat analysis network (JTAN) project.

