

# DATA MINING TOR, SOCIAL NETWORKS, OSINT WITH AIL PROJECT

E.102

CIRCL COMPUTER INCIDENT RESPONSE CENTER LUXEMBOURG

MISP PROJECT

<https://www.misp-project.org/>

APRIL 8, 2022 - VO.7



2022-04-08

Data mining Tor, social networks, OSINT with AIL  
Project

DATA MINING TOR, SOCIAL NETWORKS,  
OSINT WITH AIL PROJECT

E.102

CIRCL COMPUTER INCIDENT RESPONSE CENTER LUXEMBOURG

MISP PROJECT  
<https://www.misp-project.org/>

APRIL 8, 2022 - VO.7



# INTRODUCTION

2022-04-08

Data mining Tor, social networks, OSINT with AIL  
Project

└ Introduction

INTRODUCTION

- **Deep Web** is the part of World Wide Web not indexed or directly accessible by standard web search-engines;
- This can be content hidden from **crawlers** by requiring a specific access and this can include private social media, password-protected forums or content protected by different measures such as paywalls or specific security interface to access the information;
- A large portion of content accessible via Internet is part of the deep web<sup>1</sup>.

<sup>1</sup>also called invisible web, hidden web or non-indexed web

1. There is a huge misconception about the difference between the darknet and deep web. The differences are important because it's two different aspects which can be related to each other.

- **Deep Web** is the part of World Wide Web not indexed or directly accessible by standard web search-engines;
- This can be content hidden from **crawlers** by requiring a specific access and this can include private social media, password-protected forums or content protected by different measures such as paywalls or specific security interface to access the information;
- A large portion of content accessible via Internet is part of the deep web<sup>1</sup>.

<sup>1</sup>also called invisible web, hidden web or non-indexed web

- **Darknet** is an overlay network running on top of Internet requiring specific software to access the network and its services;
- Tor, I2P and Freenet are the most commonly used ones. Many are used for hidden services access and some for proxy access to the Internet;
- There are **legitimate use-cases** for such network but also many **illegal or criminal usage**.

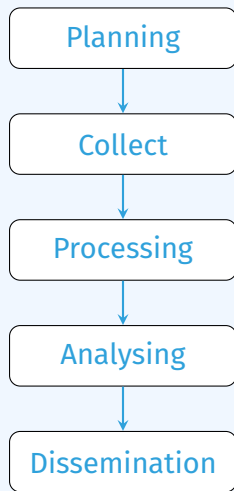
2022-04-08

## Data mining Tor, social networks, OSINT with AIL Project

### └ Introduction

### └ Concepts - darknet

- **Darknet** is an overlay network running on top of Internet requiring specific software to access the network and its services;
- Tor, I2P and Freenet are the most commonly used ones. Many are used for hidden services access and some for proxy access to the Internet;
- There are **legitimate use-cases** for such network but also many **illegal or criminal usage**.



2022-04-08

Data mining Tor, social networks, OSINT with AIL Project

└ Introduction

└ Lifecycle of collection and analysis



# COLLECTING, PROCESSING AND ANALYSING CONTENT - WEB PAGES

- Building a search engine on the web is a challenging task because:
  - ▶ it has to crawl webpages,
  - ▶ it has to make sense of **unstructured data**,
  - ▶ it has to **index** these data,
  - ▶ it has to provide a way to retrieve data and structure data (e.g. correlation).
- Doing so on Tor is even more challenging because:
  - ▶ services don't always want to be found,
  - ▶ parts of the dataset have to be discarded.
- in each case, it requires a lot of bandwidth, storage and computing power.

2022-04-08

## Data mining Tor, social networks, OSINT with AIL Project

### └ Introduction

### └ Collecting, processing and analysing content

- Building a search engine on the web is a challenging task because:
  - ▶ it has to crawl webpages,
  - ▶ it has to make sense of **unstructured data**,
  - ▶ it has to **index** these data,
  - ▶ it has to provide a way to retrieve data and structure data (e.g. correlation).
- Doing so on Tor is even more challenging because:
  - ▶ services don't always want to be found,
  - ▶ parts of the dataset have to be discarded.
- in each case, it requires a lot of bandwidth, storage and computing power.

# COLLECTING, PROCESSING AND ANALYSING CONTENT - STRUCTURED DATA

- Some data are structured and are easy to process:
  - ▶ metadata!
  - ▶ API responses.
- Some even provide cryptographic evidences:
  - ▶ authentication mechanisms between peers,
  - ▶ OpenPGP can leak a lot of metadata
    - key ids,
    - subject of email in thunderbird,
  - ▶ Bitcoin's Blockchain is public,
  - ▶ pivoting on these data with external sources yields interesting results.

- Some data are structured and are easy to process:
  - ▶ metadata!
  - ▶ API responses.
- Some even provide cryptographic evidences:
  - ▶ authentication mechanisms between peers,
  - ▶ OpenPGP can leak a lot of metadata
    - key ids,
    - subject of email in thunderbird,
  - ▶ Bitcoin's Blockchain is public,
  - ▶ pivoting on these data with external sources yields interesting results.

# AIL DESIGN OBJECTIVES

2022-04-08

Data mining Tor, social networks, OSINT with AIL Project

└ AIL design Objectives

AIL DESIGN OBJECTIVES



# OBJECTIVES OF THE SESSION

- Show how to use and extend an open source tool to monitor web pages, pastes, forums and hidden services
- Explain challenges and the design of the AIL open source framework
- Review different **collection mechanisms** and **sources**
- Learn how to create new modules
- Learn how to use, install and start AIL
- **Supporting investigation using the AIL framework** and including it in cyber threat intelligence lifecycle

2022-04-08

## Data mining Tor, social networks, OSINT with AIL Project

### └ AIL design Objectives

### └ Objectives of the session

#### OBJECTIVES OF THE SESSION

- Show how to use and extend an open source tool to monitor web pages, pastes, forums and hidden services
- Explain challenges and the design of the AIL open source framework
- Review different **collection mechanisms** and **sources**
- Learn how to create new modules
- Learn how to use, install and start AIL
- **Supporting investigation using the AIL framework** and including it in cyber threat intelligence lifecycle

# AIL FRAMEWORK

2022-04-08

Data mining Tor, social networks, OSINT with AIL  
Project

└ AIL Framework

AIL FRAMEWORK

# FROM A REQUIREMENT TO A SOLUTION: AIL FRAMEWORK

## History:

- AIL initially started as an **internship project** (2014) to evaluate the feasibility to automate the analysis of (un)structured information to find leaks.
- In 2019, AIL framework is an **open source software** in Python. The software is actively used (and maintained) by CIRCL and many organisations.
- In 2020, AIL framework is now a complete project called **ail project**<sup>2</sup>.

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

2022-04-08

## Data mining Tor, social networks, OSINT with AIL Project

### └ AIL Framework

### └ From a requirement to a solution: AIL

#### History:

- AIL initially started as an **internship project** (2014) to evaluate the feasibility to automate the analysis of (un)structured information to find leaks.
- In 2019, AIL framework is an **open source software** in Python. The software is actively used (and maintained) by CIRCL and many organisations.
- In 2020, AIL framework is now a complete project called **ail project**<sup>2</sup>.

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

# CAPABILITIES OVERVIEW

2022-04-08

Data mining Tor, social networks, OSINT with AIL Project

└ Capabilities Overview

CAPABILITIES OVERVIEW

- **Check** if mail/password/other sensitive information (terms tracked) leaked
- **Detect** reconnaissance of your infrastructure
- **Search** for leaks inside an archive
- **Monitor** and crawl websites

- Proactive investigation: leaks detection
  - ▶ List of emails and passwords
  - ▶ Leaked database
  - ▶ AWS Keys
  - ▶ Credit-cards
  - ▶ PGP private keys
  - ▶ Certificate private keys
- Feed Passive DNS or any passive collection system
- CVE and PoC of vulnerabilities most used by attackers

- Proactive investigation: leaks detection
  - ▶ List of emails and passwords
  - ▶ Leaked database
  - ▶ AWS Keys
  - ▶ Credit-cards
  - ▶ PGP private keys
  - ▶ Certificate private keys
- Feed Passive DNS or any passive collection system
- CVE and PoC of vulnerabilities most used by attackers

- Website monitoring
  - ▶ monitor booters
  - ▶ Detect encoded exploits (WebShell, malware encoded in Base64...)
  - ▶ SQL injections
- Automatic and manual submission to threat sharing and incident response platforms
  - ▶ MISP
  - ▶ TheHive
- Term/Regex/Yara monitoring for local companies/government

### └ Capabilities Overview

### └ Support CERT/CSIRTs and Law Enforcement

- Website monitoring
  - ▶ monitor booters
  - ▶ Detect encoded exploits (WebShell, malware encoded in Base64...)
  - ▶ SQL injections
- Automatic and manual submission to threat sharing and incident response platforms
  - ▶ MISP
  - ▶ TheHive
- Term/Regex/Yara monitoring for local companies/government

- Example: <https://gist.github.com/>
  - ▶ Easily storing and sharing text online
  - ▶ Used by programmers and legitimate users
    - Source code & information about configurations

### └ Capabilities Overview

### └ Sources of leaks: Paste monitoring

- Example: <https://gist.github.com/>
  - ▶ Easily storing and sharing text online
  - ▶ Used by programmers and legitimate users
    - Source code & information about configurations



- Example: <https://gist.github.com/>
  - ▶ Easily storing and sharing text online
  - ▶ Used by programmers and legitimate users
    - Source code & information about configurations
- Abused by attackers to store:
  - ▶ List of vulnerable/compromised sites
  - ▶ Software vulnerabilities (e.g. exploits)
  - ▶ Database dumps
    - User data
    - Credentials
    - Credit card details
  - ▶ More and more ...

- Example: <https://gist.github.com/>
  - ▶ Easily storing and sharing text online
  - ▶ Used by programmers and legitimate users
    - Source code & information about configurations
- Abused by attackers to store:
  - ▶ List of vulnerable/compromised sites
  - ▶ Software vulnerabilities (e.g. exploits)
  - ▶ Database dumps
    - User data
    - Credentials
    - Credit card details
  - ▶ More and more ...

# WHY SO MANY LEAKS?

- Economical interests (e.g. Adversaries promoting services)
- Ransom model (e.g. To publicly pressure the victims)
- Political motives (e.g. Adversaries showing off)
- Collaboration (e.g. Criminals need to collaborate)
- Operational infrastructure (e.g. malware exfiltrating information on a pastie website)
- Mistakes and errors

2022-04-08

## Data mining Tor, social networks, OSINT with AIL Project

### └ Capabilities Overview

#### └ Why so many leaks?

WHY SO MANY LEAKS?

- Economical interests (e.g. Adversaries promoting services)
- Ransom model (e.g. To publicly pressure the victims)
- Political motives (e.g. Adversaries showing off)
- Collaboration (e.g. Criminals need to collaborate)
- Operational infrastructure (e.g. malware exfiltrating information on a pastie website)
- Mistakes and errors

# ARE LEAKS FREQUENT?

Yes!

and we have to deal with this as a CSIRT.

- **Contacting companies or organisations** who did specific accidental leaks
- **Discussing with media** about specific case of leaks and how to make it more practical/factual for everyone
- Evaluating the economical market for cyber criminals (e.g. DDoS booters<sup>3</sup> or reselling personal information - reality versus media coverage)
- Analysing collateral effects of malware, software vulnerabilities or exfiltration

→ And it's important to detect them automatically.

<sup>3</sup><https://github.com/D4-project/>

2022-04-08

Data mining Tor, social networks, OSINT with AIL Project

└ Capabilities Overview

└ Are leaks frequent?

ARE LEAKS FREQUENT?

Yes!  
and we have to deal with this as a CSIRT.

- **Contacting companies or organisations** who did specific accidental leaks
- **Discussing with media** about specific case of leaks and how to make it more practical/factual for everyone
- Evaluating the economical market for cyber criminals (e.g. DDoS booters<sup>3</sup> or reselling personal information - reality versus media coverage)
- Analysing collateral effects of malware, software vulnerabilities or exfiltration

→ And it's important to detect them automatically.

<https://github.com/D4-project/>

## ■ Monitored paste sites: 27

- ▶ *gist.github.com*
- ▶ *ideone.com*
- ▶ ...

	2016	2017	08.2018
Collected pastes	18,565,124	19,145,300	11,591,987
Incidents	244	266	208

**Table:** Pastes collected and incident<sup>4</sup> raised by CIRCL

<sup>4</sup><http://www.circl.lu/pub/tr-46>

### └ Capabilities Overview

### └ Paste monitoring at CIRCL: Statistics

PASTE MONITORING AT CIRCL: STATISTICS

■ Monitored paste sites: 27

- ▶ gist.github.com
- ▶ ideone.com
- ▶ ...

	2016	2017	08.2018
Collected pastes	18,565,124	19,145,300	11,591,987
Incidents	244	266	208

**Table:** Pastes collected and incident<sup>4</sup> raised by CIRCL

<sup>4</sup><http://www.circl.lu/pub/tr-46>

# CURRENT CAPABILITIES

2022-04-08

Data mining Tor, social networks, OSINT with AIL Project

└─ Current capabilities

CURRENT 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 input**
- Tor Crawler (handle cookies authentication)

### └─ Current capabilities

### └─ AIL Framework: Current 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 input**
- Tor Crawler (handle cookies authentication)

# AIL FRAMEWORK: CURRENT FEATURES

- Extracting **credit cards numbers, credentials, phone numbers, ...**
- Extracting and validating potential **hostnames**
- Keeps track of **duplicates**
- Submission to threat sharing and incident response platform (**MISP** and **TheHive**)
- **Full-text indexer** to index unstructured information
- **Tagging** for classification and searches
- Terms, sets, regex and YARA **tracking and occurrences**
- Archives, files and raw **submission** from the UI
- PGP, Cryptocurrency, Decoded (Base64, ...) and username Correlation
- And many more

2022-04-08

## Data mining Tor, social networks, OSINT with AIL Project

└ Current capabilities

└ AIL Framework: Current features

AIL Framework: Current features

- Extracting **credit cards numbers, credentials, phone numbers, ...**
- Extracting and validating potential **hostnames**
- Keeps track of **duplicates**
- Submission to threat sharing and incident response platform (**MISP** and **TheHive**)
- **Full-text indexer** to index unstructured information
- **Tagging** for classification and searches
- Terms, sets, regex and YARA **tracking and occurrences**
- Archives, files and raw **submission** from the UI
- PGP, Cryptocurrency, Decoded (Base64, ...) and username Correlation
- And many more

- Search and monitor specific keywords/patterns
  - ▶ Automatic Tagging
  - ▶ Email Notifications
- Track Term
  - ▶ ddos
- Track Set
  - ▶ booter,ddos,stresser;2
- Track Regex
  - ▶ circl\.lu
- YARA rules
  - ▶ <https://github.com/ail-project/ail-yara-rules>

## Data mining Tor, social networks, OSINT with AIL Project

└─ Current capabilities

└─ Terms Tracker

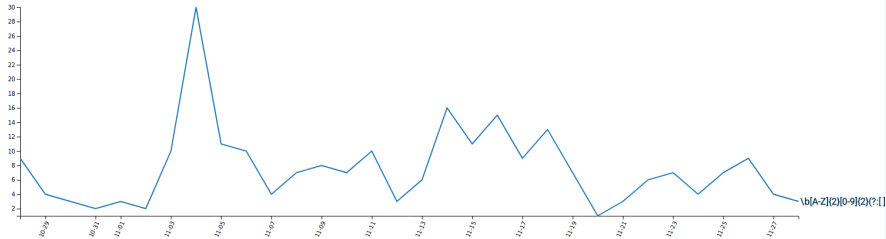
- Search and monitor specific keywords/patterns
  - ▶ Automatic Tagging
  - ▶ Email Notifications
- Track Term
  - ▶ ddos
- Track Set
  - ▶ booter,ddos,stresser;2
- Track Regex
  - ▶ circl\.lu
- YARA rules
  - ▶ <https://github.com/ail-project/ail-yara-rules>



# TERMS TRACKER

82a87a6a-88f1-4ab1-ba53-1bf15211b4b8

Type	Tracker	Date added	Level	Created by	First seen	Last seen	Tags	Email
regex	<code>\b[A-Z](2)[0-9](2)(?-[ ]?[0-9](4))(4)(?[ ]?[0-9](3))(?[ ]?[0-9](1,2))?b</code>	2019/09/12	1	admin@admin.test	2018/08/31	2019/11/28		



yyyy-mm-dd

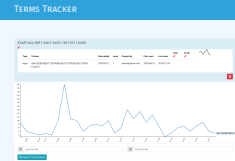
Search Tracked Items

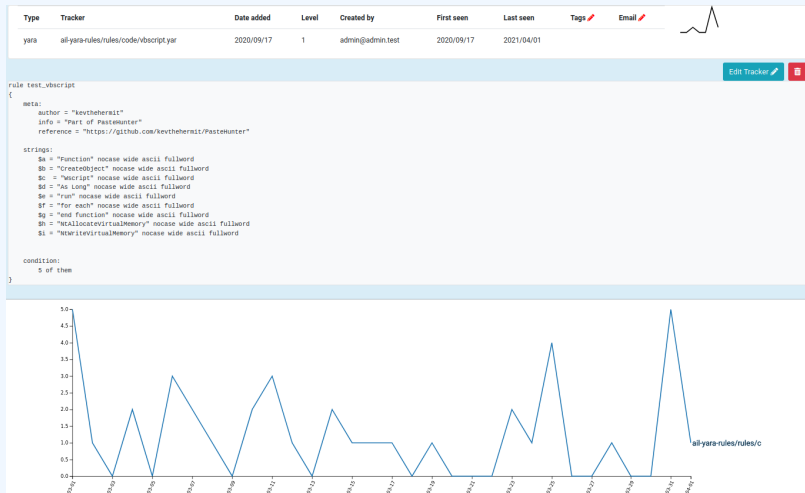
2022-04-08

Data mining Tor, social networks, OSINT with AIL Project


└─ Current capabilities


└─ Terms Tracker






## ■ Create and test your own tracker


 Tags (optional, space separated)

 E-Mails Notification (optional, space separated)

 Tracker Description (optional)

– Select a tracker type –

+ Add Tracker

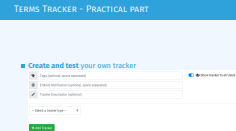
☒  Show tracker to all Users

2022-04-08

Data mining Tor, social networks, OSINT with AIL Project

└─ Current capabilities

└─ Terms Tracker - Practical part



## ■ Attacker also share informations

### ■ Recon tools detected: 94

- ▶ sqlmap
- ▶ dnscan
- ▶ whois
- ▶ msfconsole (metasploit)
- ▶ dnmap
- ▶ nmap
- ▶ ...

2022-04-08

Data mining Tor, social networks, OSINT with AIL Project

└─ Current capabilities

└─ Recon and intelligence gathering tools

- Attacker also share informations
- Recon tools detected: 94
  - ▶ sqlmap
  - ▶ dnscan
  - ▶ whois
  - ▶ msfconsole (metasploit)
  - ▶ dnmap
  - ▶ nmap
  - ▶ ...

# RECON AND INTELLIGENCE GATHERING TOOLS

```
#####
=====
Hostname      www.pabloquintanilla.cl      ISP      Wix.com Ltd.
Continent     North America              Flag
US
Country       United States              Country Code   US
Region        Unknown                    Local time    19 Nov 2019 07:59 CST
City          Unknown                    Postal Code    Unknown
IP Address    185.230.60.195              Latitude      37.751
                                           Longitude    -97.822
=====
#####
> www.pabloquintanilla.cl
Server:       38.132.106.139
Address:      38.132.106.139#53

Non-authoritative answer:
www.pabloquintanilla.cl canonical name = www192.wixdns.net.
www192.wixdns.net      canonical name = balancer.wixdns.net.
Name:   balancer.wixdns.net
Address: 185.230.60.211
>
#####
Domain name: pabloquintanilla.cl
Registrant name: SERGIO TORO
Registrant organisation:
Registrar name: NIC Chile
Registrar URL: https://www.nic.cl
Creation date: 2018-11-21 14:34:34 CLST
```

22 :34 CLST

2022-04-08

## Data mining Tor, social networks, OSINT with AIL Project

└ Current capabilities

└ Recon and intelligence gathering tools

RECON AND INTELLIGENCE GATHERING TOOLS

```
#####
=====
Hostname      www.pabloquintanilla.cl      ISP      Wix.com Ltd.
Continent     North America              Flag
US
Country       United States              Country Code   US
Region        Unknown                    Local time    19 Nov 2019 07:59 CST
City          Unknown                    Postal Code    Unknown
IP Address    185.230.60.195              Latitude      37.751
                                           Longitude    -97.822
=====
#####
> www.pabloquintanilla.cl
Server:       38.132.106.139
Address:      38.132.106.139#53

Non-authoritative answer:
www.pabloquintanilla.cl canonical name = www192.wixdns.net.
www192.wixdns.net      canonical name = balancer.wixdns.net.
Name:   balancer.wixdns.net
Address: 185.230.60.211
>
#####
Domain name: pabloquintanilla.cl
Registrant name: SERGIO TORO
Registrant organisation:
Registrar name: NIC Chile
Registrar URL: https://www.nic.cl
Creation date: 2018-11-21 14:34:34 CLST
Name server: 185.230.60.211
```

- Search for encoded strings
  - ▶ Base64
  - ▶ Hexadecimal
  - ▶ Binary
- Guess Mime-type
- Correlate paste with decoded items

└─ Current capabilities

└─ Decoder

- Search for encoded strings
  - ▶ Base64
  - ▶ Hexadecimal
  - ▶ Binary
- Guess Mime-type
- Correlate paste with decoded items

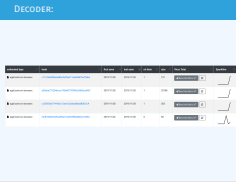
# DECODER:

2022-04-08

Data mining Tor, social networks, OSINT with AIL Project

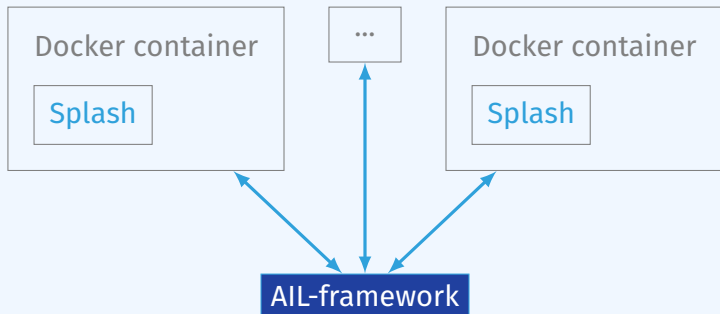
└─ Current capabilities

└─ Decoder:



estimated type	hash	first seen	last seen	nb item	size	Virus Total	Sparkline
application/x-dosexec	<a href="#">c11c2be8d9ba4e86c8effaa411aa6b867ba75abe</a>	2019/11/28	2019/11/28	1	191	<a href="#">Send this file to VT</a> <a href="#">Refresh</a>	
application/x-dosexec	<a href="#">a50cba731204ecce193b40178399a250b5ce6f67</a>	2019/11/28	2019/11/28	1	32768	<a href="#">Send this file to VT</a> <a href="#">Refresh</a>	
application/x-dosexec	<a href="#">cc5f2f0da71f443ec12ae1b3cb6ab8bad80f22c4</a>	2019/11/28	2019/11/28	1	203	<a href="#">Send this file to VT</a> <a href="#">Refresh</a>	
application/x-dosexec	<a href="#">eed67e8fa9cb9a43fea21ae653983a8e0a174f63</a>	2019/11/26	2019/11/28	6	83	<a href="#">Send this file to VT</a> <a href="#">Refresh</a>	

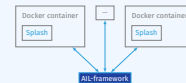
- Crawlers are used to navigate on regular website as well as .onion addresses (via automatic extraction of urls or manual submission)
- Splash ("scriptable" browser) is rendering the pages (including javascript) and produce screenshots (HAR archive too)



└ Current capabilities

└ Crawler

- Crawlers are used to navigate on regular website as well as .onion addresses (via automatic extraction of urls or manual submission)
- Splash ("scriptable" browser) is rendering the pages (including javascript) and produce screenshots (HAR archive too)





## How a domain is crawled by default

1. Fetch the first url
2. Render javascript (webkit browser)
3. Extract all urls
4. Filter url: keep all url of this domain
5. crawl next url (max depth = 1)

└─ Current capabilities

└─ Crawler

How a domain is crawled by default  
1. Fetch the first url  
2. Render javascript (webkit browser)  
3. Extract all urls  
4. Filter url: keep all url of this domain  
5. crawl next url (max depth = 1)

# 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

```
{  "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": "047ab0cd97ff5bcc77edb6a"}
```

```
{  "name": "remember_token",  "value": "12158cddd1511d74d341f23"}
```

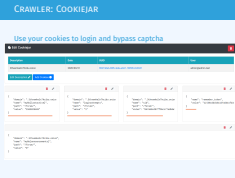
```
{  "domain": ".3thxemke2x7hcibu.onion",  "name": "mybb[announcements]",  "path": "/forum/",  "value": "0"}
```

2022-04-08


Data mining Tor, social networks, OSINT with AIL Project

└ Current capabilities

└ Crawler: Cookiejar



# CRAWLER: COOKIEJAR


**3thxmke2x7hcibu.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  Export

Decoded **1**



Screenshot **138**

Crawled Items

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

Show 10 entries Search:

Crawled Pastes

## Shere Khan

Portal Search Member List Help

Welcome back, zutepori. You last visited: 03-20-2020, 01:35 PM Log Out

You have 2 unread private messages. The most recent is from Jack3 titled **KEY FOR PRIVATE SECTIONS**

Shere Khan - Official Forum

Private Messages

Menu

user CP Home

Messages

Tracking

Your Profile

Miscellaneous

Inbox | Compose Message | Manage Folders | Empty Folders | Download Messages

1% of PM space used.



Inbox

Message Title	Sender	Date/Time Sent (sec)
<b>KEY FOR PRIVATE SECTIONS</b>	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 (Archive) Mode Mark all forums read RSS Syndication

Powered by  © 2002-2020  Group.

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

<http://3thxmke2x7hcibu.onion/forum/private.php>

2022-04-08

Data mining Tor, social networks, OSINT with AIL Project

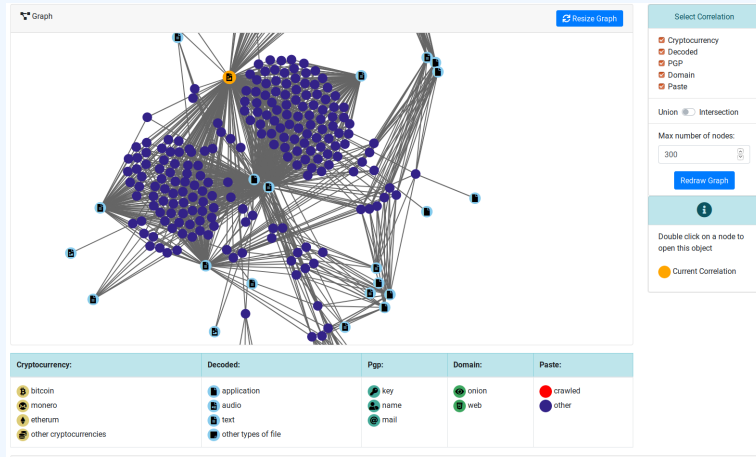
└ Current capabilities

└ Crawler: Cookiejar

CRAWLER: COOKIEJAR



# CORRELATIONS AND RELATIONSHIP



2022-04-08

Data mining Tor, social networks, OSINT with AIL Project

└ Current capabilities

└ Correlations and relationship



**LIVE DEMO!**

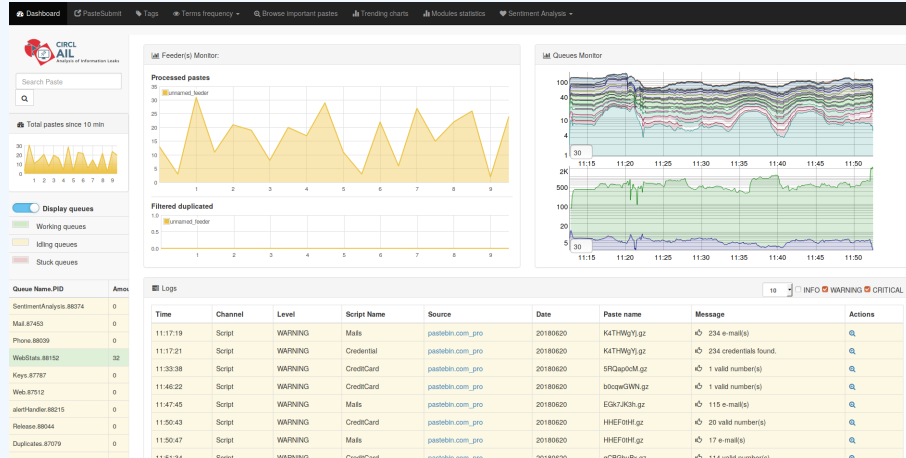
2022-04-08

Data mining Tor, social networks, OSINT with AIL  
Project

└─ Live demo!

LIVE DEMO!

# EXAMPLE: DASHBOARD

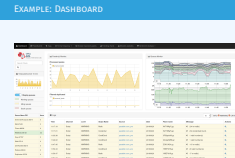


2022-04-08

Data mining Tor, social networks, OSINT with AIL Project

└ Live demo!

└ Example: Dashboard



# EXAMPLE: TEXT SEARCH

Q 1 Results for "gandcrab"

Index: 2019-05-20 - 1365.328591 Mb

Show 10 entries Search:

#	Path	Date	Size (Kb)	Action
0	<a href="#">crawled/2019/05/17/vs5e7g245s3pxjoc.onion374a1a89-4b16-4c3f-a460-4be8898da140</a> <a href="#">crawled</a> <a href="#">cve</a>	2019/05/17	15.44	<a href="#">i</a> <a href="#">Q</a>

Showing 1 to 1 of 1 entries

Previous 1 Next

Totalling 1 results related to paste content

Data mining Tor, social networks, OSINT with AIL Project

└ Live demo!

└ Example: Text search



# EXAMPLE: ITEMS METADATA (1)


infoleak:automatic-detection="phone-number"

infoleak:automatic-detection="mail"

infoleak:automatic-detection="base64"

+

Date	Source	Encoding	Language	Size (Kb)	Mime	Number of lines	Max line length
04/05/2019	pastebin.com_pro	text/plain	None	6.12	text/plain	1650	100

Create  Event

Duplicate list:

Show 

10

 entries

Search:

Hash type	Paste info	Date	Path	Action
[*]ish	Similarity: [19]%	2019-04-13	archive/pastebin.com_pro/2019/04/13/EbMVR87S.gz	<a href="#">🔗</a>
[*]ish	Similarity: [10]%	2019-04-11	archive/pastebin.com_pro/2019/04/11/2X5HRvix.gz	<a href="#">🔗</a>
[*]ish	Similarity: [23]%	2019-04-25	archive/pastebin.com_pro/2019/04/25/TS2b6M4c.gz	<a href="#">🔗</a>
[*]ish	Similarity: [14]%	2019-04-17	archive/pastebin.com_pro/2019/04/17/CuS93H7K.gz	<a href="#">🔗</a>
[*]ish	Similarity: [23]%	2019-04-20	archive/pastebin.com_pro/2019/04/20/AQd0qGVQ.gz	<a href="#">🔗</a>
[*]ish	Similarity: [20]%	2019-04-20	archive/pastebin.com_pro/2019/04/20/6DDc13b8.gz	<a href="#">🔗</a>
[*]ish	Similarity: [21]%	2019-05-05	alerts/pastebin.com_pro/2019/05/05/X8nJLzda.gz	<a href="#">🔗</a>
[*]ish	Similarity: [7]%	2019-04-13	archive/pastebin.com_pro/2019/04/13/Lyp4FVWW.gz	<a href="#">🔗</a>

Showing 1 to 8 of 8 entries

Previous 

1

 Next

Data mining Tor, social networks, OSINT with AIL Project

└ Live demo!

└ Example: Items Metadata (1)

Item	Source	Encoding	Language	Size (Kb)	Mime	Number of lines	Max line length
1	pastebin.com_pro	text/plain	None	6.12	text/plain	1650	100
Duplicate list:							
Hash type	Paste info	Date	Path	Action			
[*]ish	Similarity: [19]%	2019-04-13	archive/pastebin.com_pro/2019/04/13/EbMVR87S.gz	<a href="#">🔗</a>			
[*]ish	Similarity: [10]%	2019-04-11	archive/pastebin.com_pro/2019/04/11/2X5HRvix.gz	<a href="#">🔗</a>			
[*]ish	Similarity: [23]%	2019-04-25	archive/pastebin.com_pro/2019/04/25/TS2b6M4c.gz	<a href="#">🔗</a>			
[*]ish	Similarity: [14]%	2019-04-17	archive/pastebin.com_pro/2019/04/17/CuS93H7K.gz	<a href="#">🔗</a>			
[*]ish	Similarity: [23]%	2019-04-20	archive/pastebin.com_pro/2019/04/20/AQd0qGVQ.gz	<a href="#">🔗</a>			
[*]ish	Similarity: [20]%	2019-04-20	archive/pastebin.com_pro/2019/04/20/6DDc13b8.gz	<a href="#">🔗</a>			
[*]ish	Similarity: [21]%	2019-05-05	alerts/pastebin.com_pro/2019/05/05/X8nJLzda.gz	<a href="#">🔗</a>			
[*]ish	Similarity: [7]%	2019-04-13	archive/pastebin.com_pro/2019/04/13/Lyp4FVWW.gz	<a href="#">🔗</a>			





## EXAMPLE: ITEMS METADATA (2)

### Hash files:

Show  entries

Search:

estimated type	hash	saved_path	Virus Total
 application/octet-stream	3975f058bb0d445b60c10a11f1a5d88e19e4fa84 (1)	HASHS/application/octet-stream /39/3975f058bb0d445b60c10a11f1a5d88e19e4fa84	<a href="#">Send this file to VT</a> 
 application/octet-stream	fed93c1753270fc849a4db37027b569cdd9a6108 (1)	HASHS/application/octet-stream /fe/fed93c1753270fc849a4db37027b569cdd9a6108	<a href="#">Send this file to VT</a> 

Showing 1 to 2 of 2 entries

Previous **1** Next

2022-04-08

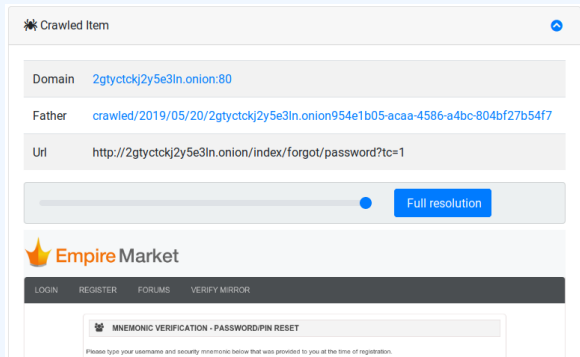
Data mining Tor, social networks, OSINT with AIL Project

└ Live demo!

└ Example: Items Metadata (2)



## EXAMPLE: ITEMS METADATA (3)



2022-04-08

Data mining Tor, social networks, OSINT with AIL Project

└ Live demo!

└ Example: Items Metadata (3)



# EXAMPLE: BROWSING CONTENT

## Content:

```
http://members2.mofosnetwork.com/access/login/
somoextremos:buddy1990
brazzers_glenn:cocklick
brazzers61:braves01

http://members.naughtyamerica.com/index.php?m=login
gernblanston:3unc2352
Janhuss141200:310575
igetalliwant:1377zeph
pwilks89:mon22key
Bman1551:hockey

MoFos IKnowThatGirl PublicPickUps
http://members2.mofos.com
Chrismagg40884:loganm40
brando1:zzbrando1
aacoen:1q2w3e4r
1rstunkle23:my8self

BraZZers
http://ma.brazzers.com
gcjensen:gcj21pva
skycsc17:rbcndnd
```

```
#####
>| Get Daily Update Fresh Porn Password Here |<

=> http://www.erq.io/4mF1
```

2022-04-08

Data mining Tor, social networks, OSINT with AIL Project

└ Live demo!

└ Example: Browsing content

EXAMPLE: BROWSING CONTENT

## Content:

```
#####
>| Get Daily Update Fresh Porn Password Here |<

=> http://www.erq.io/4mF1
```

# EXAMPLE: BROWSING CONTENT

## Content:

Over 50000+ custom hacked xxx passwords by us! Thousands of free xxx passwords to the hottest paysites!

#####

>| Get Fresh New Premium XXX Site Password Here |<

=>    http://www.erq.io/4mF1

#####

http://ddfnetwork.com/home.html

eu172936:hCS8gKh

UecwB6zs:159X0\$r#6K78FuU

http://pornxn.stiffia.com/user/login

feldwWek8939:R0bluJ8XtB

dabudka:17891789

brajits:brajits1

http://members.pornstarplatinum.com/sblogin/login.php/

gigiriveracom:xxxjay

jayx123:xxxjay69

http://members.vividceleb.com/

Rufio99:fairhaven

ScHiFRvi:102091

Chaos84:HOLE5244

Riptor795:blade7

Dom180:harkonnen

GaggedUK:a1k0chan

http://www.ariellaferrera.com/

2022-04-08

Data mining Tor, social networks, OSINT with AIL Project

└─ Live demo!

└─ Example: Browsing content

EXAMPLE: BROWSING CONTENT

## Content

Over 50000+ custom hacked xxx passwords by us! Thousands of free xxx passwords to the hottest paysites!

#####

>| Get Fresh New Premium XXX Site Password Here |<

=>    http://www.erq.io/4mF1

#####

http://ddfnetwork.com/home.html

eu172936:hCS8gKh

UecwB6zs:159X0\$r#6K78FuU

http://pornxn.stiffia.com/user/login

feldwWek8939:R0bluJ8XtB

dabudka:17891789

brajits:brajits1

http://members.pornstarplatinum.com/sblogin/login.php/

gigiriveracom:xxxjay

jayx123:xxxjay69

http://members.vividceleb.com/

Rufio99:fairhaven

ScHiFRvi:102091

Chaos84:HOLE5244

Riptor795:blade7

Dom180:harkonnen

GaggedUK:a1k0chan

http://www.ariellaferrera.com/

# EXAMPLE: SEARCH BY TAGS

Search Tags by date range :

2019-05-19

2019-05-21

infoleak.automatic-detection="cve" infoleak.automatic-detection="bitcoin-address"

Search Tags

Show

10

Search:

entries

Date	Path	# of lines	Action
2019/05/19	archive/pastebin.com_pro/2019/05/19/ej67tQ4b.gz <div>cve bitcoin-address</div>	71	<div></div>
2019/05/21	archive/pastebin.com_pro/2019/05/21/vM2SwyTe.gz <div>cve bitcoin-address</div>	69	<div></div>
2019/05/21	archive/pastebin.com_pro/2019/05/21/rsnHnp5L.gz <div>cve bitcoin-address</div>	71	<div></div>

Showing 1 to 3 of 3 entries

Previous

1

Next

2022-04-08

Data mining Tor, social networks, OSINT with AIL Project

└─ Live demo!

└─ Example: Search by tags

EXAMPLE: SEARCH BY TAGS



# MISP

2022-04-08

Data mining Tor, social networks, OSINT with AIL  
Project  
└─ MISP

MISP

- **Tagging** is a simple way to attach a classification to an event or an attribute.
- **Classification must be globally used to be efficient.**
- Provide a set of already defined classifications modeling estimative language
- Taxonomies are implemented in a simple JSON format <sup>5</sup>.
- Can be easily cherry-picked or extended

<sup>5</sup><https://github.com/MISP/misp-taxonomies>

- **Tagging** is a simple way to attach a classification to an event or an attribute.
- **Classification must be globally used to be efficient.**
- Provide a set of already defined classifications modeling estimative language
- Taxonomies are implemented in a simple JSON format <sup>5</sup>.
- Can be easily cherry-picked or extended

- **infoleak**: Information classified as being potential leak.
- **estimative-language**: Describe quality and credibility of underlying sources, data, and methodologies.
- **admiralty-scale**: Rank the reliability of a source and the credibility of an information
- **fpr**<sup>6</sup>: Evaluate the degree of identifiability of personal data and the types of pseudonymous data, de-identified data and anonymous data.

<sup>6</sup>Future of Privacy Forum

- **infoleak**: Information classified as being potential leak.
- **estimative-language**: Describe quality and credibility of underlying sources, data, and methodologies.
- **admiralty-scale**: Rank the reliability of a source and the credibility of an information
- **fpr**<sup>6</sup>: Evaluate the degree of identifiability of personal data and the types of pseudonymous data, de-identified data and anonymous data.

<sup>6</sup>Future of Privacy Forum



- **tor**: Describe Tor network infrastructure.
- **dark-web**: Criminal motivation on the dark web.
- **copine-scale**<sup>7</sup>: Categorise the severity of images of child sex abuse.

<sup>7</sup>Combating Paedophile Information Networks in Europe

- **tor**: Describe Tor network infrastructure.
- **dark-web**: Criminal motivation on the dark web.
- **copine-scale**<sup>7</sup>: Categorise the severity of images of child sex abuse.

<sup>7</sup>Combating Paedophile Information Networks in Europe

# THREAT SHARING AND INCIDENT RESPONSE PLATFORMS



**Goal:** submission to threat sharing and incident response platforms.

2022-04-08

Data mining Tor, social networks, OSINT with AIL Project

└ MISP

└ threat sharing and incident response



Goal: submission to threat sharing and incident response platforms.



1. Use infoleak taxonomy<sup>8</sup>
2. Add your own tags
3. Export AIL objects to MISP core format
4. Download it or Create a MISP Event<sup>9</sup>

<sup>8</sup><https://www.misp-project.org/taxonomies.html>

<sup>9</sup><https://www.misp-standard.org/rfc/misp-standard-core.txt>

Data mining Tor, social networks, OSINT with AIL  
Project  
└ MISP

└ threat sharing and incident response




1. Use infoleak taxonomy<sup>8</sup>
2. Add your own tags
3. Export AIL objects to MISP core format
4. Download it or Create a MISP Event<sup>9</sup>

<sup>8</sup><https://www.misp-project.org/taxonomies.html>  
<sup>9</sup><https://www.misp-standard.org/rfc/misp-standard-core.txt>

# MISP EXPORT

1Gt545E48EPsyTC8voKQDCFpTkwiuXduw :

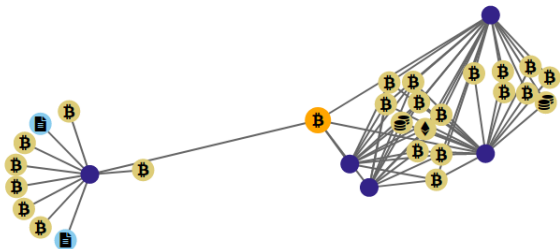
Object type	type	First seen	Last seen	Nb seen
cryptocurrency	 bitcoin	2020/01/17	2020/02/20	5

Expand Bitcoin address

Graph

Resize Graph

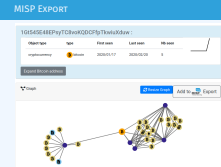
Add to  Export



2022-04-08

Data mining Tor, social networks, OSINT with AIL  
Project  
└─ MISP

└─ MISP Export



nttfj36sp47cw2yecop572zjvjeazgazieunllouudplzqt2m  
5h465yd.onion :



First Seen Last Check Ports

2020/02/19 2020/02/19 [80]

infoleak:automatic-detection="onion"



Last Origin: [crawled/2020/02/19/dark.failc126d32a-3ed1-468f-ba24-f2e5956f4035](#)

🔍 Show Domain Correlations 4

Add to  Export

🖼 Screenshot 4



Hide

[LOGIN](#)
[REGISTER](#)
[FORUMS](#)
[VER](#)

Login

LOGIN TO EMPIRE MARI
 

Welcome to Empire Market! Please log  
Registrations are free and open to every

Username  
 Password

What's th

Login

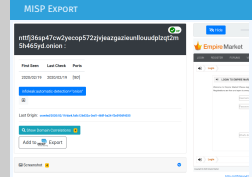
Copyright © 2020 Empire Market


2022-04-08

Data mining Tor, social networks, OSINT with AIL  
Project

└─ MISP

└─ MISP Export




**MISP Exporter**

Select a list of objects to export

Object Type	Object ID		Lvl		
Object type...			0		+
Object type...	1Gt545E48EPsyTC8voKQDCfpTkwuXduw	✓	1		
Domain	nttfj36sp47cw2yecop572zvjvjeazgzieunllouudplzqt2m5h465yd.onion	✓	0		

JSON Export ☒ Export to MISP Instance

Distribution: Your organisation only

Threat Level: Medium

Analysis: Initial

Event Info: Quick Event Description or Tracking Info

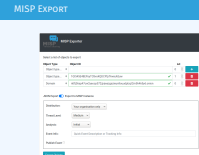
Publish Event ☐

Export Objects

## Data mining Tor, social networks, OSINT with AIL Project

└ MISP


└ MISP Export



# AUTOMATIC SUBMISSION ON TAGS

MISP Auto Event Creation


Enabled



✖ Disable Event Creation

The hive auto export

Disabled



✔ Enable Alert Creation

Metadata : 6 / 25

Show 5 entries Search:

Whitelist	Tag
<input checked="" type="checkbox"/>	infoleak:automatic-detection="api-key"
<input checked="" type="checkbox"/>	infoleak:automatic-detection="aws-key"
<input checked="" type="checkbox"/>	infoleak:automatic-detection="base64"
<input type="checkbox"/>	infoleak:automatic-detection="bitcoin-address"
<input type="checkbox"/>	infoleak:automatic-detection="bitcoin-private-key"

Showing 1 to 5 of 25 entries

Previous 1 2 3 4 5

Next

Metadata : 23 / 25

Show 5 entries Search:

Whitelist	Tag
<input checked="" type="checkbox"/>	infoleak:automatic-detection="api-key"
<input checked="" type="checkbox"/>	infoleak:automatic-detection="aws-key"
<input checked="" type="checkbox"/>	infoleak:automatic-detection="base64"
<input checked="" type="checkbox"/>	infoleak:automatic-detection="bitcoin-address"
<input checked="" type="checkbox"/>	infoleak:automatic-detection="bitcoin-private-key"

Showing 1 to 5 of 25 entries

Previous 1 2 3 4 5

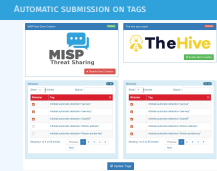
Next

✔ Update Tags

2022-04-08

Data mining Tor, social networks, OSINT with AIL Project  
└ MISP

└ Automatic submission on tags



API

2022-04-08

Data mining Tor, social networks, OSINT with AIL  
Project  
└ API

API



AIL exposes a ReST API which can be used to interact with the back-end<sup>10</sup>.

```
curl https://127.0.0.1:7000/api/v1/get/item/default
  --header "Authorization: iHc1_ChZxj1aXmiFiF1m"
  -H "Content-Type: application/json"
  --data @input.json -X POST
```

- AIL API is currently covering 60% of the functionality of back-end.

---

<sup>10</sup><https://github.com/ail-project/ail-framework/blob/master/doc/README.md>

# SETTING UP THE FRAMEWORK

2022-04-08

Data mining Tor, social networks, OSINT with AIL Project

└─ Setting up the framework

SETTING UP THE FRAMEWORK

## Setting up AIL-Framework from source

```
1 git clone  
  https://github.com/ail-project/ail-framework.git  
2 cd AIL-framework  
3 ./installing_deps.sh
```

# FEEDING THE FRAMEWORK

2022-04-08

Data mining Tor, social networks, OSINT with AIL Project

└ Feeding the framework

FEEDING THE FRAMEWORK

There are different way to feed AIL with data:

1. Setup *pystemon* and use the custom feeder
  - ▶ *pystemon* will collect items for you
2. Use the new JSON Feeder (twitter)
3. Feed your own data using the API or the `import_dir.py` script
4. Feed your own file/text using the UI (Submit section)

## Data mining Tor, social networks, OSINT with AIL Project

### └ Feeding the framework

### └ Feeding AIL

There are different way to feed AIL with data:

1. Setup *pystemon* and use the custom feeder
  - ▶ *pystemon* will collect items for you
2. Use the new JSON Feeder (twitter)
3. Feed your own data using the API or the `import_dir.py` script
4. Feed your own file/text using the UI (Submit section)

# VIA THE UI (1)

Files submission

Submit a file

Browse...

No file selected.

Archive Password

Optional

Tags :

Select Tags

Taxonomie Selection ▼

Select Tags

Galaxy Selection ▼

Submit this paste

2022-04-08

Data mining Tor, social networks, OSINT with AIL Project

└ Feeding the framework

└ Via the UI (1)

VIA THE UI (1)

Submit a file

Browse...

No file selected.

Archive Password

Optional

Tags :

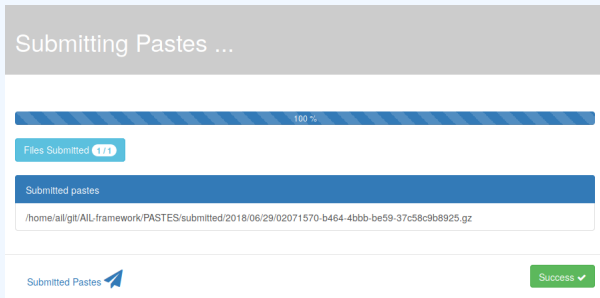
Select Tags

Taxonomie Selection ▼

Select Tags

Galaxy Selection ▼

Submit this paste



## api/v1/import/item

```
1 {  
2   "type": "text",  
3   "tags": [  
4     "infoleak:analyst-detection=\"private-key\""  
5   ],  
6   "text": "text to import"  
7 }
```

```
api/v1/import/item  
{  
  "type": "text",  
  "tags": [  
    "infoleak:analyst-detection=\"private-key\""  
  ],  
  "text": "text to import"  
}
```



/!\ requirements:

- Each file to be fed must be of a reasonable size:
  - ▶ ~ 3 Mb / file is already large
  - ▶ This is because some modules are doing regex matching
  - ▶ If you want to feed a large file, better split it in multiple ones

- Each file to be fed must be of a reasonable size:
  - ▶ ~ 3 Mb / file is already large
  - ▶ This is because some modules are doing regex matching
  - ▶ If you want to feed a large file, better split it in multiple ones

# FEEDING AIL WITH YOUR OWN DATA - import\_dir.py (2)

1. Check your local configuration `configs/core.cfg`
  - ▶ In the file `configs/core.cfg`,
  - ▶ Add `127.0.0.1:5556` in `ZMQ_Global`
  - ▶ (should already be set by default)
2. Launch `import_dir.py` with the directory you want to import
  - ▶ `import_dir.py -d dir_path`

2022-04-08

Data mining Tor, social networks, OSINT with AIL Project

└ Feeding the framework

└ Feeding AIL with your own data -

FEEDING AIL WITH YOUR OWN DATA -  
import\_dir.py (2)

1. Check your local configuration `configs/core.cfg`
  - ▶ In the file `configs/core.cfg`,
  - ▶ Add `127.0.0.1:5556` in `ZMQ_Global`
  - ▶ (should already be set by default)
2. Launch `import_dir.py` with the directory you want to import
  - ▶ `import_dir.py -d dir_path`