DATA MINING TOR, SOCIAL NETWORKS, OSINT WITH AIL PROJECT

E.102

CIRCL COMPUTER INCIDENT RESPONSE CENTER LUXEMBOURG

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

October 27, 2022 - V0.7





Threat Sharing



-27

2022

Data mining Tor, social networks, OSINT with AIL Project

DATA MINING TOR, SOCIAL NETWORKS, OSINT WITH AIL PROJECT E:102 CIRCL COMPUTER INCOMENT RESPONSE CONTRE LIXED PROFES



Data mining Tor, social networks, OSINT with AIL Project

INTRODUCTION

INTRODUCTION

CONCEPTS - DEEP WEB



- This can be content hidden from crawlers by requiring a specific access and this can includes 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 web¹.

```
Data mining Tor, social networks, OSINT with AIL
Ŀ,
  Project
      Introduction
2022
```

Deep Web is the part of World Wide Web not indexed

-Concepts - Deep 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.

¹also called invisible web, hidden web or non-indexed web

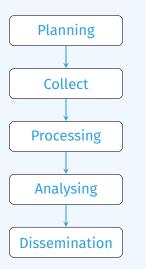
CONCEPTS - DARKNET

- 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.

```
Data mining Tor, social networks, OSINT with AIL
Project
Introduction
```

- Darknet is an overlay network running on top of Interne 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.

LIFECYCLE OF COLLECTION AND ANALYSIS



Data mining Tor, social networks, OSINT with AIL Project

└─Introduction

-27

2022-10

Lifecycle of collection and analysis



COLLECTING, PROCESSING AND ANALYSING CONTENT - WEB PAGES

- Data mining Tor, social networks, OSINT with AIL
 - Introduction

2022

Collecting, processing and analysing content

COLLECTING, PROCESSING AND ANALYSING CONTENT NEB PAGES

- Building a search engine on the web is a challenging task because:
- it has to crawl webpages,
- it has to to make sense of unstructured data,
 it has to index these 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 an

- Building a search engine on the web is a challenging task because:
 - it has to crawl webpages,
 - it has to 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

Data mining Tor, social networks, OSINT with AIL Project

Introduction

-27

-10

2022

```
Collecting, processing and analysing content
```

COLLECTING, PROCESSING AND ANALYSING CONTENT STRUCTURED DATA

- Some data are structured and are easy to process metadatal
- API responses.
- Some even provide cryptographic evidences
- authentication mechanisms between peer
 OpenGPG can leak a lot of metadata
 - Openation can wake a rot or instantial
 key ids,
 subject of enabling threaderhind
 - subject of email in thunderbird,
 Bitcoin's Blockchain is public
- Bitcoin's Biockchain is public,
 pivoting on these data with external sources yiel

- Some data are structured and are easy to process:
 - metadata!
 - ► API responses.
- Some even provide cryptographic evidences:
 - authentication mechanisms between peers,
 - OpenGPG 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

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

```
Data mining Tor, social networks, OSINT with AIL
Project
AIL design Objectives
```

-Objectives of the session

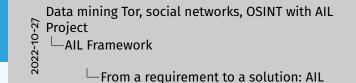
OBJECTIVES OF THE SESSIO

Show how to use and extend an open source tool to monitor web pages, pastes, forums and hidden services of papalan challenges and the design of the AL open source framework. In Review different collection mechanisms and sources a Learn how to create new modules Learn how to create new modules Learn how to create new modules (Learn how to create new modules and the second second second second second second collecting investigation using the AL framework and including in in order threat intelligence lifecycle **AIL FRAMEWORK**

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

AIL FRAMEWORK

FROM A REQUIREMENT TO A SOLUTION: AIL FRAMEWORK



FROM A REQUIREMENT TO A SOLUTION: AIL FRAMEWORK

- AlL initially started as an internship project (2014) to evaluate the feasibility to automate the analysis of (un)structured information to find leaks.
- In 2019, AlL framework is an open source software in Python. The software is actively used (and maintained) by CIDCL and many activities.
- In 2020, AlL framework is now a complete project called ail project².

https://github.com/ail-project

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**².

CAPABILITIES OVERVIEW

Data mining Tor, social networks, OSINT with AIL Project Capabilities Overview

CAPABILITIES OVERVIEW

COMMON USAGE

Data mining Tor, social networks, OSINT with AIL Project Capabilities Overview ISAGE

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

- 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

SUPPORT CERT/CSIRTS AND LAW ENFORCEMENT ACTIVITIES

Data mining Tor, social networks, OSINT with AIL Project

Capabilities Overview

-27

2022-10

-Support CERT/CSIRTs and Law Enforcement

SUPPORT CERT/CSIRTS AND LAW ENFORCEMENT ACTIVITIES

- Proactive investigation: leaks detection
- List of emails and passwords
- Leaked dat
 AWS Keys
- AWS Keys
 Credit-card
- PGP private keys
- Certificate private keys
- Feed Passive DNS or any passive collection system
 CVE and PoC of subprabilities most used by attacker

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

SUPPORT CERT/CSIRTS AND LAW ENFORCEMENT ACTIVITIES

Data mining Tor, social networks, OSINT with AIL Project

└─Capabilities Overview

-27

2022-10

-Support CERT/CSIRTs and Law Enforcement

SUPPORT CERT/CSIRTS AND LAW ENFORCEMENT ACTIVITIES

- Website monitoring
- monitor booters
 Detect encoded exploits (WebShell, malware encoded in
- Base64...)
- Automatic and manual submission to threat sharing and
 - ent response platforms
- MISP
 TheHive
- Term/Regex/Yara monitoring for loca

- 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

SOURCES OF LEAKS: PASTE MONITORING

- Data mining Tor, social networks, OSINT with AIL Project
- └─Capabilities Overview

-27

-9

2022-

└─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 configurat

Example: https://gist.github.com/

- **•** Easily storing and sharing text online
- Used by programmers and legitimate users
 - \rightarrow Source code & information about configurations

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
- Abused by attackers to store:
 - List of vulnerable/compromised sites
 - Software vulnerabilities (e.g. exploits)
 - Database dumps
 - ightarrow User data
 - $\rightarrow \text{Credentials}$
 - \rightarrow Credit card details
 - More and more ...

- Data mining Tor, social networks, OSINT with AIL Project
- └─Capabilities Overview

-21

-9

2022-

└─Sources of leaks: Paste monitoring

- SOURCES OF LEAKS: PASTE MONITORIN
 - Example: https://gist.github.com/ Example: https://gist.github.com/
- Used by programmers and legitimate users
 Source code & information about configuration
- Source code & information about configu
 Abused by attackers to store:
- List of vulnerable/compromised siter
- Software vulnerabilities (e.g. explo
 Database dumps

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

Data mining Tor, social networks, OSINT with AIL Project Capabilities Overview WHY SO MANY LEAKS

 Economical interests (e.g. Adversaries promoting services) Bansom model (e.g. To publicy pressure the vicinim) Political motives (e.g. Adversaries showing off) Collaboration (e.g. Criminals med to collaborate) Operational infrastructure (e.g. malavae oditrating Information on a pastie website) Mitables 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 booters³ or reselling personal information reality versus media coverage)
- Analysing collateral effects of malware, software vulnerabilities or exfiltration
 - \rightarrow And it's important to detect them automatically.

```
Data mining Tor, social networks, OSINT with AIL
Project
Capabilities Overview
```

ARE LEAKS FREQUE

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 booters⁵ or reselling personal information - reality versus media coverage)
- Analysing collateral effects of malware, software vulnerabilities or exfiltration

And it's important to detect them autor s://github.com/Di-project/

³https://github.com/D4-project/

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 incident⁴ raised by CIRCL

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

Capabilities Overview

-27

0

2022

Paste monitoring at CIRCL: Statistics

| Monitored paste sit | 05: 33 | |
|-------------------------------------|--------|---------|
| aist.aithub.com | | |
| ideone com | | |
| F | | |
| | | |
| | | |
| | 2016 | |
| Collected pastes | | 08.2018 |

http://www.circl.lu/pub/tr-46

⁴http://www.circl.lu/pub/tr-46

CURRENT CAPABILITIES

Data mining Tor, social networks, OSINT with AIL Project Current capabilities

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)

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

— Current capabilities
```

AIL Framework: Current capabilities

- Extending AIL to add a new analysis module can be done in go 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 Multiple concurrent data input
- Tor Crawler (handle cookies authentica

-21

-9

2022-

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 occurences
- Archives, files and raw **submission** from the UI
- PGP, Cryptocurrency, Decoded (Base64, ...) and username Correlation
- And many more

Data mining Tor, social networks, OSINT with AIL Project —Current capabilities

'n

2022-10

AIL Framework: Current features

AIL FRAMEWORK: CURRENT FEATU

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

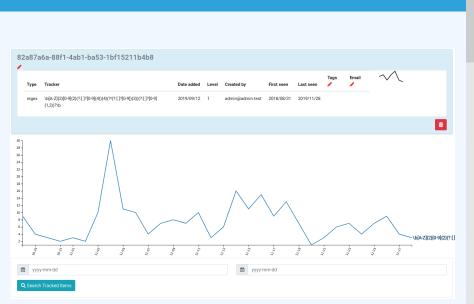
YARA rules

https://github.com/ail-project/ail-yara-rules

Data mining Tor, social networks, OSINT with AIL Project Current capabilities

MS TRACKER

Search and monitor specific keywords/patterns
Automatic Tagging
Track Term
Adda
Track Search
Bootecddos,thresser;
Controls
Contr



Data mining Tor, social networks, OSINT with AIL Project Current capabilities

└─Terms Tracker

2022-10-27



YARA TRACKER

| Туре | Tracker | Date added | Level | Created by | First seen | Last seen | Tags 🥖 | Email 🧪 | . / |
|--|--|------------|-------|------------------|------------|------------------|--|---------|--------------------|
| yara | ail-yara-rules/rules/code/vbscript.yar | 2020/09/17 | 1 | admin@admin.test | 2020/09/17 | 2021/04/01 | | | 20 |
| | | | | | | | | | Edit Tracker 🖋 🥫 |
| info refe strings: sa = sb = sc : sd = se = sf = sg = sh = | er "Yerthöhersi" * Twei di Astenuer* ene "Nitjui/Dib.dox/Architersi/Pistek "rontion" nocae side acii fullard "orestedget:" nocae side acii fullard "read" nocae side acii fullard "read" nocae side acii fullard "read" nocae side acii fullard "read" nocae side acii fullard "maltionary nocae side acii fullard "maltionary nocae side acii fullard | llword | | | | | | | |
| | | | | | | -tr ₀ | 1973 - 1975 - 19 | - Reb | aliyas ridesindewo |

Data mining Tor, social networks, OSINT with AIL Project Current capabilities

└─YARA Tracker

2022-10-27



TERMS TRACKER - PRACTICAL PART

Data mining Tor, social networks, OSINT with AIL Project

└─Current capabilities

-27

2022-10

└─Terms Tracker - Practical part

| | Engo (spitorai, spore separated) | C deinormake ball |
|---|------------------------------------|-------------------|
| ۰ | Estais technological, pace age and | |
| 1 | | |

Create and test your own tracker

| ٠ | Tags (optional, space separated) | 💽 🚢 Show tracker to all Users |
|------|--|-------------------------------|
| @ | E-Mails Notification (optional, space separated) | |
| ø | Tracker Description (optional) | |
| | | |
| - Se | elect a tracker type - 🔶 | |
| | | |
| - A | dd Tracker | |
| TA | | |
| | | |

RECON AND INTELLIGENCE GATHERING TOOLS

Attacker also share informations

Recon tools detected: 94

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

Data mining Tor, social networks, OSINT with AIL Project

Current capabilities

-27

2022-10

Recon and intelligence gathering tools

- Attacker also share information
 Recon tools detected: 94
- sqlmap
 dnscan
- whois
 msfronsole (metaso
- dnmap
 nman

RECON AND INTELLIGENCE GATHERING TOOLS

| Hostname | www.pabloquinta | nilla.cl | | | ISP | Wix.com | Ltd. |
|-----------------|------------------|----------|----------|----------|-----------|----------|-----------|
| Continent US | North America | | Flag | | | | |
| Country | United States | | Country | Code | US | | |
| Region Unknown | | Local t | ime | 19 Nov | 2019 07: | 59 CST | |
| City Unknown | | Postal | Code | Unknowr | 1 I | | |
| IP Address | 185.230.60.195 | | Latitud | е | 37.751 | | |
| | Longitu | de | -97.822 | | | | |
| | | | | | | | |
| ****** | ***** | ######## | ######## | 4####### | *######## | ######## | ********* |
| > www.pabloquin | | | | | | | |
| Server: | 38.132.106.139 | | | | | | |
| Address: | 38.132.106.139# | 53 | | | | | |
| Non-authoritati | ve answer: | | | | | | |
| www.pabloquinta | nilla.cl canonic | al name | = www192 | .wixdns. | net. | | |
| www192.wixdns.n | et canonic | al name | = balanc | er.wixdr | ns.net. | | |
| Name: balance | r.wixdns.net | | | | | | |
| Address: 185.23 | 0.60.211 | | | | | | |
| > | | | | | | | |
| | | ######## | ######## | | | ######## | |
| Domain name: pa | bloquintanilla.c | 1 | | | | | |
| Registrant name | : SERGIO TORO | | | | | | |
| Registrant orga | nisation: | | | | | | |
| Pogistrar name: | NIC Chile | | | | | | |
| Negistiai name. | | | | | | | |
| | https://www.nic. | cl | | | | | |

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

```
└─Current capabilities
```

-27

9

2022-

```
Recon and intelligence gathering tools
```

| | UNLINE DESCENT | | | |
|--|-------------------|----------|--|--|
| | | 1184 | | |
| | | 1000 | | |
| | | | | |
| | 285, 238, 69, 285 | | | |
| | 6 mg/m | - 10.412 | | |
| | | | | |
| | | | | |
| | | | | |
| | | | | |
| | | | | |
| | | | | |
| | | | | |
| | | | | |
| | | | | |
| | Codeting cast | | | |
| | | | | |
| | | | | |
| | | | | |
| | | | | |
| | | | | |
| | | | | |
| | | | | |

DECODER

Search for encoded strings

- ► Base64
- ► Hexadecimal
- Binary
- Guess Mime-type
- Correlate paste with decoded items

Data mining Tor, social networks, OSINT with AIL Project Current capabilities

DECODER

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

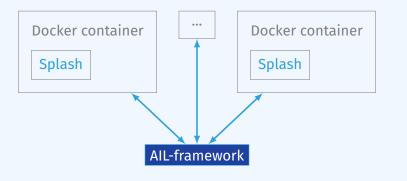
DECODER:

Data mining Tor, social networks, OSINT with AIL Project Current capabilities

| estimated type | hash 🍴 | first seen 🛛 🗍 | last seen 🛛 🕸 | nb item 📋 | size 📋 | Virus Total | Sparkline 🌐 |
|-----------------------|--|----------------|---------------|-----------|--------|----------------------|---------------|
| application/x-dosexec | c11c2be8d9ba4e86c8effaa411aa6b867ba75abe | 2019/11/28 | 2019/11/28 | 1 | 191 | Send this file to VT | |
| application/x-dosexec | a50cba731204ecce193b40178399a250b5ce6f67 | 2019/11/28 | 2019/11/28 | 1 | 32768 | Send this file to VT | |
| application/x-dosexec | cc5f2f0da711443ec12ae1b3cb6ab8bad80f22c4 | 2019/11/28 | 2019/11/28 | 1 | 203 | Send this file to VT | / |
| application/x-dosexec | eed67e8fa9cb9a43fea21ae653983a8e0a174f63 | 2019/11/26 | 2019/11/28 | 6 | 83 | Send this file to VT | $_ \! \land$ |

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 rending the pages (including javascript) and produce screenshots (HAR archive too)



Data mining Tor, social networks, OSINT with AIL Project Current capabilities

CRAWLER



CRAWLER

```
Data mining Tor, social networks, OSINT with AIL
Project
Current capabilities
```

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 | | | | ē |
|--|--|----------------------|---|---|
| Description | Date | UUID | | User |
| 3thxemke2x7hcibu.onion | 2020/03/31 | 90674deb-38fb-4eba-a | 1661-18899ccb3841 | admin@admin.test |
| Edit Description 🖋 Add Cookies 😌 | | | | |
| i / | | i / | i / | i / |
| <pre>{ "domain": ".3thxemke2x7hcibu.onior "name": "mybb[Lastactive]", "path": "/forum/", "value": "1583829465" }</pre> | <pre>{ "domain": ".3thxemk "name": "loginattem "path": "/forum/", "value": "i" }</pre> | | { "domain": ".3thxemke2x7hcibu.onior "name": "sid", "path": "/forum/", "value": "047ab0cd97ff5bcc77edb0a: } | <pre>{ "name": "remember_token", "value": "12 S8cddd1511d74d341f23+ }</pre> |
| | | | | i / |
| <pre>{ "domain": ".3thxemke2x7hcibu.onion", "name": "myb0[announcements]", "path": "/forum/", "value": "0" }</pre> | | | | |

Data mining Tor, social networks, OSINT with AIL Project

└─Current capabilities

-27

2022-10

Crawler: Cookiejar

| Previation | - | *** | | - |
|---------------------|--------|-----|------------------|-------------|
| Pharties Tells over | 88.871 | | | strepsto of |
| Gast Station | , 30.3 | | ar teachertarras | |
| | | | | |

CRAWLER: COOKIEJAR

| | | Shere Kh | an | | |
|--|------------------------------------|---|--|---------------------------------------|---------------------------------|
| First Seen Last Check Ports | | | | | iearch 🙎 Member List 📱 H |
| | | Welcome back, zulopori. | You last visited: 03-20-2020, 01:35 PM Log (| out → New Posts View Today's Posts | |
| 2020/03/09 2020/03/30 ['80'] | | Se User CP | View | New Posts View Today's Posts | Private Hessages (Uncend 2, Tot |
| | | | w have 2 unread private messages. The most recen | t is from Jok3 10ed KEY FOR PRIVATE | SECTIONS |
| nfoleak:automatic-detection="onion" infoleak:automatic-detection="onion" infoleak:automatic-detection="onion | etection="base64" | Shere Khan - Official Forum Private Messages | | | |
| Ŧ | | Henu | | | Luc Luci |
| 5 | | Clier CP Home | Inbox Compose Message Manage Folders Empty | rolders Download Hessages | 1% of PM space us |
| | | | Inbox | Enter Keywords | Search PHs (Advanced Se |
| nual | | Compose | Message Title | Sender | Date/Time Sent [asc |
| | | L Dobax | KEY FOR PRIVATE SECTIONS | Jok3 | 3 hours ag |
| Show Domain Correlations 139 | | L Drafts L Trash Can | Verification | Jok3 | 03-09-2020, 11:55 AM |
| | | Tracking | | Move To Inbox 💌 | or Delete the selected mes |
| Add to MISP Export | | 🔀 Edit Folders | | | amp to Felders Inbox 💌 |
| Thosa Sharing | | Your Profile | | | |
| | | L grange Password | | | |
| Decoded 1 | | L Dornge Aveter L / Change Signature | | | |
| Decoded 1 | | @ Edit Options | | | |
| | | Miscellaneous 🚥 | | | |
| Screenshot 138 | | Group Memberships Buddy/Sprore List | | | |
| Screensnot 138 | • | // Nanage Attachments | | | |
| | | Saved Drafts | | | |
| | | Subscribed Threads | | | |
| rawled Items Dat | te: 2020/03/23 - 13:10.40 PORT: 80 | | | | |
| | | Forum Team Contact Us S | here Khan - Hacking group Return to Top Life (| Archiva) Hode Mark all forums read | RSS Syndication |
| 10 ¢ entries | Search: | Powered By HyBB, @ 2002-2021 | Ny55 Group | | Current time: 03-23-2020, 0 |

Data mining Tor, social networks, OSINT with AIL Project

└─Current capabilities

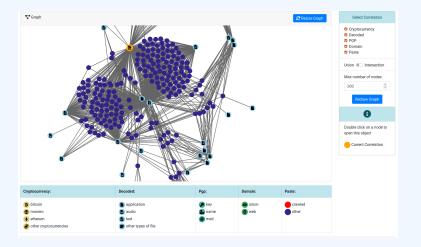
2022-10-27

Crawler: Cookiejar

| re Klass |
|----------|
| |
| |
| 2221 |
| |
| |
| |
| |
| |
| |
| |

54

CORRELATIONS AND RELATIONSHIP



- Data mining Tor, social networks, OSINT with AIL Project
- └─Current capabilities

2022-10-27

Correlations and relationship



Data mining Tor, social networks, OSINT with AIL Project Live demo!

LIVE DEMO!

LIVE DEMO!

EXAMPLE: DASHBOARD

Data mining Tor, social networks, OSINT with AIL Project

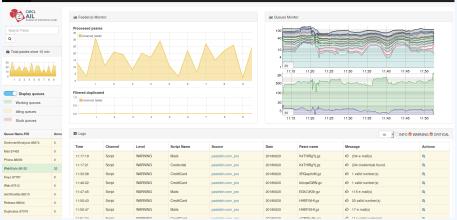
Live demo!

2022-10-27

Example: Dashboard

| Annal Mil | 10.21 | 14 PC | | - - | | | | |
|--|-------|-------|------|------------|------|---|----------|--|
| A 10 | | - | | | - | | | |
| | | 7 | | | 1 | Y | | |
| in the local division of the local division | 1.00 | 1.0 | | | | | | |
| | | | | | | | | |
| | | | | | | | | |
| | 1 | | | | | | | |
| | | | | | | | | |
| | | | | | | | | |
| hard and the state of the state | | | | | | | | |
| har and Manager San Par | | | | | | | | |
| Anna ann Anna ann Anna ann Anna ann | | | - | | | | | |
| har and Manager San Par | | | | | | | P Carlos | |

🚓 Dashboard 🤨 PasteSubmit 🔖 Tags 👁 Terms frequency - 🔍 Browse important pastes 🖬 Trending charts 🖬 Modules statistics 🎔 Sentiment Analysis -



EXAMPLE: TEXT SEARCH

| Q 1 Resul | ts for "gandcrab" | | | |
|-----------|---|------------|-----------|-----------|
| Index: | 2019-05-20 - 1365.328591 Mb | | Search: | |
| # 41 | Path 1 | Date 11 | Size (Kb) | Action 1 |
| 0 | crawled/2019/05/17/vs5e7g245s3pxjoc.onion374a1a89-4b16-4c31-a460-4be8898da140 | 2019/05/17 | 15.44 | છ |
| Showing 1 | to 1 of 1 entries | | Previo | us 1 Next |
| Totalling | 1 results related to paste content | | | |

Data mining Tor, social networks, OSINT with AIL Project Live demo!

Example: Text search

| et an i tana nanenar lada 🛛 🖷 | | | | | |
|-------------------------------|-----------|--|----------|--|--|
| | | | hest. | | |
| | | | 510 (10) | | Anise |
| | Designed. | | 19.44 | | |
| | een. | ensi i be Antonio Congrega contra do atracto de astronomio de Estato | | In the second se | ente Derita de la constante de |

EXAMPLE: ITEMS METADATA (1)

| inioleak:autom | auc-detect | ion="phone-number" | intolea | k:automatic-de | stection= | man. | infoleak:automatic-d | etection= pase64" | + | | | |
|----------------|------------|--------------------|---------|----------------|-----------|-----------|---------------------------|-------------------|-----------------|------------|--------|--|
| Date | Sourc | 0 | End | oding | Langua | ge | Size (Kb) | Mime | Number of lines | Max line I | ength | |
| 04/05/2019 | pastet | bin.com_pro | text | /plain | None | | 6.12 | text/plain | 1650 | 100 | | |
| Create MISP E | | | | | | | | | | | | |
| Duplicate list | les | Paste info | | Date | | Path | | | | Search: | Action | |
| ['tish'] | - | Similarity: [19]% | | 2019-04-13 | | | astebin.com pro/2019/04/1 | 2/EbM/D97S az | | | | |
| ['tish'] | | Similarity: [10]% | | 2019-04-11 | | | astebin.com pro/2019/04/1 | | | | | |
| ('tish') | | Similarity: [23]% | | 2019-04-25 | | | astebin.com pro/2019/04/2 | | | | | |
| ('tish') | | Similarity: [14]% | | 2019-04-17 | | | astebin.com_pro/2019/04/1 | | | | | |
| ['tlsh'] | | Similarity: [23]% | | 2019-04-20 | | archive/p | astebin.com_pro/2019/04/2 | 20/AQd0qGVQ.gz | | | | |
| ['tish'] | | Similarity: [20]% | | 2019-04-20 | | archive/p | astebin.com_pro/2019/04/2 | 20/6DDc13b8.gz | | | | |
| ['tish'] | | Similarity: [21]% | | 2019-05-05 | | alerts/pa | tebin.com_pro/2019/05/05 | /X8nJLzda.gz | | | | |
| ('tish') | | Similarity: [7]% | | 2019-04-13 | | archive/p | astebin.com_pro/2019/04/1 | 13/Lyp4FVWW.gz | | | | |
| | | | | | | | | | | | | |

Data mining Tor, social networks, OSINT with AIL Project Live demo!

Example: Items Metadata (1)

| | | ******** | ***** | 1100.0010 | | • | |
|---------------|--|----------------|----------------|-----------|---|-------------------|--|
| - | - | Service . | ing the second | - | - | Reduction at land | |
| | | | | | | | |
| Orginate | ter w | | | | | | |
| | | | | | | | |
| | - | - (m. 1994) | | | | | |
| | | | | | | | |
| 1111 | ar | | | | | | |
| 10 | Arristanti de la composición d | | | | | | |
| 1 1 1 1 1 1 1 | 1. berne 1. | | | | | | |
| TELEFE | 10 10 10 10 10 10 10 10 10 10 10 10 10 1 | | | | | | |
| 1 2 2 2 2 2 2 | 1. berne 1. | | | | | | |

-27

2022-10

Example: Items Metadata (2)

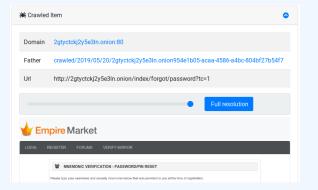
Data mining Tor, social networks, OSINT with AIL Project Live demo!

Example: Items Metadata (2)

| Hash files: Nov 1 James | | | 141 |
|----------------------------|--------------------------------|--|---|
| minuted too | | sand just | dea line |
| C salaterator dical | Statement of the second second | With additional data and a strategy of the second s | at the second |
| C aphatariatet alvan | MICCOMMENCEMENT | McCapital and and And CCCC Contend Classific D | a factoria de |
| Round La coloure | - | | Panet |
| | | | |

| Hash files: Show 5 • entries | | Se | arch: |
|---------------------------------|--|--|----------------------|
| estimated type | hash 🕌 | saved_path | Virus Total |
| application/octet- stream | 3975f058bb0d445b60c10a11f1a5d88e19e4fa84 (1) | HASHS/application/octet-stream /39/3975f058bb0d445b60c10a11f1a5d88e19e4fa84 | Send this file to VT |
| application/octet- stream | fed93c1753270fc849a4db37027b569cdd9a6108 (1) | HASHS/application/octet-stream /le/led93c1753270fc849a4db37027b569cdd9a6108 | Send this file to VT |
| Showing 1 to 2 of 2 entries | S | | Previous 1 Next |

EXAMPLE: ITEMS METADATA (3)



- Data mining Tor, social networks, OSINT with AIL Project └─Live demo!
 - Example: Items Metadata (3)

| lanas igospylalasia K Heli saka 2015 2012 algospylalasia K Heli kalagospylalasia katalogi pasaritika (kalagospylalasia) |
|---|
| |
| D1 Marchaeleybelmanan/marchaeleyatanane/%in/ |
| |
| • Fit was for |

EXAMPLE: BROWSING CONTENT

Content:

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

http://members.naughtyamerica.com/index.php?m=login gernblanston:3unc2352 Janhuss141200:310675 igetalliwant:1377zeph pwilks89:mno22key Bman1551:hockey

MoFos IKnowThatGirl PublicPickUps http://members2.mofos.com Chrismagg40884:10ganm48 brando1:zzbrando1 aacoen:1q2W3e4r irstunkle23:my8self

BraZZers http://ma.brazzers.com gcjensen:gcj21pva skycsc17:rbcdnd

>| Get Daily Update Fresh Porn Password Here |<

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

Data mining Tor, social networks, OSINT with AIL Project Live demo!

Example: Browsing content

| Content: | |
|--------------------------|--|
| http://www.st.adunters.c | who are right |
| | |
| | |
| Max meters and the state | |
| | |
| | |
| | |
| personal substrates | |
| | |
| | |
| | |
| | |
| | |
| LANSING MADE | |
| | |
| | |
| | |
| | |
| | A fail but to being the first free formers from in |
| | |
| | |

-27

2022-10

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:hCSBgKh UecwB6zs:159X0\$!r#6K78FuU

http://pornxn.stiffia.com/user/login feldwWek8939:RObluJ8XtB 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:alk8chan

http://www.ariellaferrera.com/

Data mining Tor, social networks, OSINT with AIL Project Live demo!

Example: Browsing content

-27

2022-10

| feer 2000 so in halfed on passers by all Teacher means the second second second second second second r) for first has frequencies the fact framework for (r) | of free our parameter in the initial |
|---|--------------------------------------|
| ry lan front has freedur this take feature from pr | |
| | |
| R02.000.03.00.001 | |
| | |
| https://withubach.com/home.html | |
| NUCLER NORMAL REPORT OF A DESCRIPTION | |
| Nue Aprenie al Planaerine Orgin | |
| | |
| | |
| ICA: A MARKET AND A STREET AND | |
| | |
| The shares we have an | |
| | |
| | |
| | |
| | |
| | |

EXAMPLE: SEARCH BY TAGS

| Search Tags by d | ate range : | | | | |
|--|--|--------|------------|------------|------------|
| 2019-05-19 | | | 2019-05-21 | | |
| infoleak automatic-detecti Q Search Tags | on="cve" \times [infoleak.automatic-detection="bitcoin-address" \times | | | | * |
| Show 10 ¢ entries | | Searc | 1: | | |
| Date | Path | | | # of lines | Action |
| 2019/05/19 | archive/pastebin.com_pro/2019/05/19/ej67tQ4 | lb.gz | | 71 | 0 Q |
| 2019/05/21 | archive/pastebin.com_pro/2019/05/21/vM2Sw cvo bitcoin-address | yTe.gz | | 69 | 0 @ |
| 2019/05/21 | archive/pastebin.com_pro/2019/05/21/rsnHnp: cve_bitcoin-address | 5L.gz | | 71 | 6 Q |
| Showing 1 to 3 of 3 entri | es Previous | 1 | Next | | |

Data mining Tor, social networks, OSINT with AIL Project Live demo!

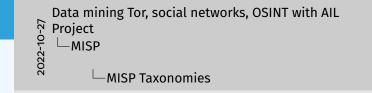
-Example: Search by tags

| 8 1111 | n | 8 1000 | | |
|------------|-------------------------|-----------------------|---------------|-----|
| | Concernent Inconcernent | | | |
| O here for | | | | |
| | | less. | | |
| | | | | |
| | | | | |
| - | M | | 4 ations | |
| - | han . | and the second second | f allow Th | _ |
| in . | | a anna ing tao ya | | • • |



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

MISP TAXONOMIES



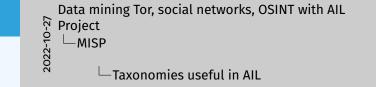
MISP TAXONOMIES

 Tagging is a simple way to attach a classification to an event or anatribude. Bobably used to be efficient. Provide a set of already defined classifications modeling estimative language Taxonomiss are implemented in a simple JSON format³. Can be easily cherry-picked or extended

https://github.com/MISP/misp-taxonomies

- **Tagging** is a simple way to attach a classification to an event or anattribute.
- 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** ⁵.
- Can be easily cherry-picked or extended

TAXONOMIES USEFUL IN AIL

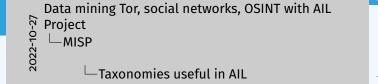


- infoleal: Information classified as being potential leak.
 estimative-language: Describe quality and credibility of underlying sources, data, and methodologies.
 admiraty-scale: Rank the reliability of a source and the credibility of an information
- fpf⁶: Evaluate the degree of identifiability of personal data and the types of pseudonymous data, de-identified data and anonymous data.

⁴Future 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
- **fpf**⁶: Evaluate the degree of identifiability of personal data and the types of pseudonymous data, de-identified data and anonymous data.

TAXONOMIES USEFUL IN AIL



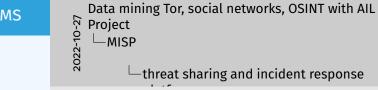
tor: Describe Tor network infrastructure.
 dark-web: Criminal motivation on the dark web.
 copine-scale?: Categorise the severity of images of child sex
 abuse.

Combating Paedophile Information Networks in Europe

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

⁷Combating Paedophile Information Networks in Europe

THREAT SHARING AND INCIDENT RESPONSE PLATFORMS







Goal: submission to threat sharing and incident response platforms.



Goal: submission to threat sharing and incident response platforms.

THREAT SHARING AND INCIDENT RESPONSE PLATFORMS



- 1. Use infoleak taxonomy⁸
- 2. Add your own tags
- 3. Export AIL objects to MISP core format
- 4. Download it or Create a MISP Event⁹

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

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



Use infoleak taxonomy^a
 Add your own tags
 Export AlL objects to MISP core format
 Download it or Create a MISP Event⁹

⁴https://www.misp-project.org/taxonomies.html ⁹https://www.misp-standard.org/rfc/misp-standard-core.tx

MISP Export

| 1G | IGt545E48EPsyTC8voKQDCFfpTkwiuXduw : | | | | | |
|-----------------------------------|---|------------|------------|--|---------------------------------------|--|
| | Object type | type | First seen | Last seen | Nb seen | |
| cryptocurrency B bitcoin 2 | | 2020/01/17 | 2020/02/20 | 5 | | |
| Ex | pand Bitcoin address | | | | | |
| Sr Graph Add to ∰ Export | | | | | | |
| | B B B B B B B B B B B B B B B B B B B | B | B | BB BB BB BB BB BB BB | B B B B B B B B B B B B B B B B B B B | |

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

-MISP Export



MISP Export

nttfj36sp47cw2yecop572zjvjeazgazieunllouudplzqt2m 5h465yd.onion :

| First Seen | Last Check | Ports | | | |
|----------------------------|--------------------|------------------|--|--|--|
| 2020/02/19 | 2020/02/19 | ['80'] | | | |
| infoleak:autor | matic-detection | ="onion" | | | |
| Ŧ | | | | | |
| Last Origin: cra | wled/2020/02/19/da | rk.failc126d | | | |
| Q Show Domain Correlations | | | | | |
| Q Show Dom | ain Correlation | s <mark>4</mark> | | | |
| Q Show Dom Add to | ain Correlations | s <mark>4</mark> | | | |
| | | s <mark>4</mark> | | | |
| Add to Miss | | s <mark>4</mark> | | | |

| 🕅 Hide | 2022-10-27 |
|--|------------|
| | 2 |
| Login Login TO EMPIRE MARI | |
| Welcome to Empire Markett Please log Registrations are the and open to every Usernam | |
| | |
| What's th | |
| Copyright 0 2020 Empire Market | |

 \bigcirc

Data mining Tor, social networks, OSINT with AIL Project

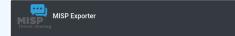
| .L | | |
|----|--|--|
| SP | | |

-MI

```
└─MISP Export
```

| | 👉 Empire Mark |
|---|---------------------------------------|
| Find Seen Laul Clanik Paris | 100 0010 AA |
| 2006/02/19 2020/03.18 (90) | < 1.00 To 0 |
| Make waarena ankar oner | Anno San Ini |
| 8 | |
| at train weatoons was a case on an up or server | |
| Q these bonain Consistons | |
| Add to and Export | |
| | • • • • • |
| PSenamhat 📕 | · · · · · · · · · · · · · · · · · · · |

MISP EXPORT



Select a list of objects to export

| Object Type | Object ID | Lv | rl | |
|----------------|--|----|----|---|
| Object type + | | 0 | < | + |
| Object type \$ | 1Gt545E48EPsyTC8voKQDCFfpTkwiuXduw ✓ | 1 | * | m |
| Domain \$ | nttfj36sp47cw2yecop572zjvjeazgazieunllouudplzqt2m5h465yd.onion | 0 | < | m |

JSON Export 💽 Export to MISP Instance

| | Distribution: | Your organisation only |
|---|---------------|--|
| | Threat Level: | Medium ~ |
| | Analysis: | Initial v |
| | Event Info: | Quick Event Description or Tracking Info |
| | Publish Event | |
| l | | |

Export Objects

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

-MISP Export

| New York Company | | |
|------------------|--------------------------------|--|
| The law | No.12 | |
| Machan. 4 | | |
| | | |
| Juna 4 | RDs/Voral/Secondostofed-Molenn | |
| Antonio C | The participants of | |
| Instant | Melan + | |
| Antipula | India - | |
| | historylawistic a fasting blu | |
| Sec. 16. | | |

AUTOMATIC SUBMISSION ON TAGS

| | MISP Threat Sharing | | |
|----------------------|--|----------------------|--|
| letadata : Show 5 | entries Search: | Metadata : Show 5 | entries Search: |
| Whitelist | Tag 👫 | Whitelist | Tag 👫 |
| | infoleak:automatic-detection="api-key" | | infoleak:automatic-detection="api-key" |
| | infoleak:automatic-detection="aws-key" | | infoleak:automatic-detection="aws-key" |
| | infoleak:automatic-detection="base64" | | infoleak:automatic-detection="base64" |
| 0 | infoleak:automatic-detection="bitcoin-address" | | infoleak:automatic-detection="bitcoin-address" |
| 0 | infoleak:automatic-detection="bitcoin-private-key" | | infoleak:automatic-detection="bitcoin-private-key" |
| Showing 1 to | o 5 of 25 entries Previous 1 2 3 4 5 Next | Showing 1 to 5 | i of 25 entries Previous 1 2 3 4 5 Next |

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

Automatic submission on tags

| | MI | SP st Sharing | 8 | TheH |
|-------------|-------------|----------------------------|--------------|--------------------------|
| interest in | teen | - | and a second | |
| - | 1.14 | | | 1× |
| | | | | |
| | allow solve | de antides (se les) | | Management and the local |
| | 1584 Aller | ale debutes lead if | | Vehicle Protection Table |
| | | an anno fear amar | | |
| • | | | | |
| • • | | ala akadan 'akan yeshi ku' | | |

-27

2022-10



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

AIL exposes a ReST API which can be used to interact with the back-end¹⁰.

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

¹⁰https://github.com/ail-project/ail-framework/blob/ master/doc/README.md Data mining Tor, social networks, OSINT with AIL Project API

AlL exposes a ReST API which can be used to interact with the back-end**.

curl https://127.0.0.1:7000/api/v1/get/item/default --header "Authorization: iHc1_ChZxjtaXmiFiFt -H "Content-Type: application/json" --data @input.json -X POST

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

"https://github.com/ail-project/ail-framework/blob/ master/doc/README.md

Data mining Tor, social networks, OSINT with AIL Project Setting up the framework

SETTING UP THE FRAMEWORK

SETTING UP THE FRAMEWORK

SETTING UP AIL-FRAMEWORK FROM SOURCE

Data mining Tor, social networks, OSINT with AIL Project

└─Setting up the framework

-27

-9

2022

Setting up AIL-Framework from source



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

Data mining Tor, social networks, OSINT with AIL Project Feeding the framework

FEEDING THE FRAMEWORK

FEEDING THE FRAMEWORK

Data mining Tor, social networks, OSINT with AIL Project Feeding the framework DING AIL

1. Setup pystemon and use the custom feede

4. Feed your own file/text using the UI (Submit section)

- 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)

Submit this paste

| Files submission | | Tags : | |
|--------------------------|--|-----------------------|---|
| Submit a file | | | ¥ |
| Browse No file selected. | | Taxonomie Selection 🗸 | |
| Archive Password | | | - |
| Optionnal | | | * |
| | | Galaxy Selection 🗸 | |
| | | | _ |

```
Data mining Tor, social networks, OSINT with AIL
Project
Feeding the framework
```

```
└─Via the UI (1)
```



VIA

| THE UI (2) | Data mining Tor, social networks, OSINT with AIL ? Project |
|---|---|
| | Project Feeding the framework |
| | |
| Submitting Pastes | |
| Files Submitted (T) Submitted pastes | |
| /home/all/git/AlL/ramework/PASTES/submitted/2018/06/29/02/071570-b464-4bbb-be59-37c58c5be925.gz | |

Submitting Pastus ... Nacional CD

1.00 A

51

FEEDING AIL WITH YOUR OWN DATA - API

Data mining Tor, social networks, OSINT with AIL Project

Feeding the framework

-27

9

2022-

Feeding AIL with your own data - API



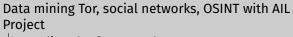
api/v1/import/item

2

6 7

"type": "text",
"tags": [
 "infoleak:analyst-detection=\"private-key\""
],
"text": "text to import"

FEEDING AIL WITH YOUR OWN DATA import_dir.py(1)



└─Feeding the framework

-27

2022-10

```
Feeding AIL with your own data -
```

FEEDING ALL WITH YOUR OWN DATA import_dir.py(1)

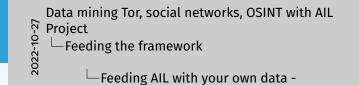
/!\ requirements

- Each file to be fed must be of a reasonable size:
- ~ 3 Hb / file is already large
 This is because some modules are doing recey mat
- If you want to feed a large file, better split it in multiple on

I\ requirements:

- Each file to be fed must be of a reasonable size:
 - \blacktriangleright ~ 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)



FEEDING AIL WITH YOUR OWN DATA import_dir.py(2)

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

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 de directory you want to
 import
 - import_dir.py -d dir_path