

AIL Framework for Analysis of Information Leaks

From a CSIRT use-case towards a generic analysis open source software



CIRCL
Computer Incident
Response Center
Luxembourg

Team CIRCL - *TLP:WHITE*

info@circl.lu

FIRST 2017

Leaks and CSIRT day-to-day operations

- Analysing and notifying about information leaks can be **time consuming** (e.g. national/sectoral CSIRT level)
- Notification can be challenging (e.g. what kind of trusted channels do you have to communicate with a victim?)
- When leaks are publicly known, **interaction with media/press** can be significant
- Analysis of mixed structured and **unstructured data** from untrusted sources (e.g. fake and duplicate leaks)

A source of leaks: Paste monitoring

- Example: <http://pastebin.com/>
 - Easily storing and sharing text online
 - Used by programmers and legitimate users
 - Source code & configuration information
- Abused by attackers to store:
 - List of vulnerable/compromised sites
 - Software vulnerability (e.g. exploits)
 - Database dumps
 - User data
 - Credentials (3rd party)
 - Credit card details
 - ... more and more ...

Paste monitoring at CIRCL: Key numbers

- Monitored paste sites: 27
- Keywords - Search terms: 420
- Keywords - Constituency related: 90
- Time for one ticket: 5 min - 1 hour

Table : Key numbers for 2016

Pastes 2016	Jan	Feb	Mar	Apr	Mai	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Fetches pastes	1 439 453	1 537 186	1 719 646	1 622 674	1 595 881	1 561 700	1 422 628	1 443 938	1 519 026	1 581 793	1 656 985	1 464 214
Keywords hits	5394	4407	4072	11 455	4722	4158	4083	3796	4235	3970	4155	4350
Constituency hits	1792	1402	741	1273	1146	795	598	644	717	953	736	643
Security related (TR-46)	30	22	28	19	15	13	16	8	13	22	38	28
Incidents & investigations	65	55	76	44	31	36	40	21	39	59	104	79

Paste monitoring: Statistics

Table : Statistics for 2016

Pastes 2016	Monthly average	Total
Fetches pastes	1 547 094	18 565 124
Keywords hits	4900	58 797
Constituency hits	953	11 440
Security related (TR-46)	21	252
Incidents & investigations	54	649

Paste monitoring: TR-46 approach

<https://www.circl.lu/pub/tr-46>



[About](#) | [Team](#) | [News](#) | [Services](#) | [Training](#) | [Publications](#) | [Projects](#) | [Contact](#)

TR-46 - Information Leaks Affecting Luxembourg and Recommendations

Overview

Information leak: the publication (or trusted announcement of possession) of stolen or otherwise acquired digital information like user profiles, credentials or other digital assets.

Information leaks have happened many times in the recent past. Sometimes, the number of affected people is quite small like in the leak of a customer database of a small web shop, where we probably would try to contact the few affected individuals or their employer's IT department. But most of the time we face leaks that contain several million people's private information.

From our experience as a CERT, it is difficult to inform individuals about the actual leak that happened. Too high is the suspicion the actual warning could be a phishing, and hence it is ignored. Testing services ("Is my email part of the leak?") have legal implications and are also problematic from a security perspective.

This document is a new approach to deal with the mass of information leaks. It is our intention to demonstrate the associated risks and suggest appropriate reactions of users of the service that leaked the information by listing the service of an information leak and showing the number of affected users in Luxembourg - as far as we know them.

TR-46 is an always-updated document. All new information leaks are mentioned here, for the

TR-46 - Information Leaks Affecting Luxembourg and Recommendations

[↑ Back to Publications and Presentations](#)

[Overview](#)

[How do I know if a service was affected?](#)

[Is CIRCL also informing me directly / my ISP / my company?](#)

[Vendor reactions](#)

[What are the risks of my information being stolen?](#)

[What should I do if the service I'm using...](#)

Paste monitoring: TR-46 approach

- How to deal with numerous requests from press/media or potential victims. The TR-46 document includes:
 - Risks with stolen email addresses
 - Risks with stolen (hashed) passwords
 - How to mitigate the risks
 - How to prevent collateral damage
 - How do we find leaks
 - **Reference of leaks** (with the number of affected users in CIRCL's constituency)
- We don't provide any form for validation of email/credential leaks. This can be conflictual with general security awareness (e.g. entering email/credentials on unknown websites).

Paste monitoring: TR-46 approach

338904	2016-11-30	unknown	y	4	73	USER ID, PASSWORD, PHONE NUMBER, RECOVERY/ALTERNATIVE EMAIL, LOCATION
339220	2016-12-01	In relation with cardio & fitness	y	31	10952	email address, password clear
341894	2016-12-04	In relation with poster & posterfuchs	y	1	3211	email address, password clear
341994	2016-12-04	www.golfersfriend.co.za	y	6	9319	username, email, password hashed, salt
344326	2016-12-08	unknown	y	1	1258	email address, password clear
344816	2016-12-09	unknown	y	1	24	email address, password clear
346932	2016-12-13	unknown	y	1	87	email address, password clear
349931	2016-12-17	In relation with tunesoman.com	y	1	7374	email address, password clear
350106	2016-12-18	In relation with Motor, Car, Mini	y	2	5661	email address, password clear
350392	2016-12-19	www.1394store.com	y	21	1349	email address, password clear
352791	2016-12-23	unknown	y	2	1289	email address, password clear
352924	2016-12-24	unknown	y	6	3734	email address, password hashed, password clear
353000	2016-12-24	seaofliveshop.com	y	13	2268	email address, password clear
353067	2016-12-25	In relation with gabon, acjaho	y	1	1543	email address, password clear
353961	2016-12-28	skillab.it	y	1	1410	user name, password hashed, email address
354507	2016-12-30	Mom-, Mommy- social community related	y	367	106187	email address, password clear
354802	2016-12-30	www.shoesontheweb.com	y	29	1863	email address, password clear
354821	2016-12-30	unknown	y	2	6910	email address, password clear
355785	2017-01-03	www.deezer.com	y	1	728	email address, password clear
355879	2017-01-03	Minecraft related	y	1	2812	email address, password clear
356313	2017-01-04	Netflix related	y	1	101	email address, password clear
357168	2017-01-07	unknown	y	1	1128	email address, password clear
357648	2017-01-09	bunkerindex.com	n	2	3985	email address
359040	2017-01-13	pile44.com, piles44.com	y	26	17472	email address, password clear
359306	2017-01-15	ludygames.com	y	3	2549	id, nom, pass, mail, passmd5, description

AIL Framework

- AIL initially started as an internship project (2014) to evaluate the feasibility to automate the analysis of (un)structured information to find leaks.
- In 2017, AIL framework is an open source software in Python. The software is actively used (and maintained)¹ by CIRCL.
- 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.

¹To follow our mantra: to "eat our own dogfood"

AIL

Q 1 Results for "B35nGGBp"

Show entries

Search:

#	Path	Date	Size (Kb)	Action
1	/home/adulau/git/AIL-framework/PASTES/archive/pastebin.com_pro/2017/01/20/B35nGGBp.gz	2017/01/20	5.8	i Q

Showing 1 to 1 of 1 entries

Previous **1** Next

Totalling 0 results related to paste content

AIL

Date	Source	Encoding	Language	Size (Kb)	Mime	Number of lines	Max line length
20/01/2017	pastebin.com_pro	text/plain	('en', 1.0)	5.8	text/plain	510	336

Duplicate list:

Show entries Search:

Hash type	Paste info	Date	Path
tlsh	Similarity: 93%	2017-01-12	/home/adulau/git/AIL-framework/PASTES/archive/pastebin.com_pro/2017/01/12/WeizLQUx.gz
tlsh	Similarity: 93%	2017-01-17	/home/adulau/git/AIL-framework/PASTES/archive/pastebin.com_pro/2017/01/17/Xqbx62vU.gz
tlsh	Similarity: 93%	2017-01-10	/home/adulau/git/AIL-framework/PASTES/archive/pastebin.com_pro/2017/01/10/iyfet4UM.gz
tlsh	Similarity: 92%	2017-01-14	/home/adulau/git/AIL-framework/PASTES/archive/pastebin.com_pro/2017/01/14/G7AB7q1m.gz
tlsh	Similarity: 92%	No date available	/home/adulau/git/AIL-framework/PASTES/archive/pastebin.com_pro/2016/12/31/CpDdkKbU.gz

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  
igetaliwant:1377zeph  
pwilks89:mon22key  
Bman1551:hockey
```

```
MoFos IKnowThatGir1 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
```

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:R0bluJ8XtB

dabudka:17891789

brajits:brajits1

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

gigiriveracom:xxxjay

jayx123:xxxjay69

http://members.vividceleb.com/

Rufio99:fairhaven

ScHiFRv1:102091

Chaos84:HOLE5244

Riptor795:blade7

Domi80:harkonnen

GaggedUK:a1k0chan

http: [REDACTED]

Browse important pastes

[Credentials](#)
[Credit cards](#)
[SQL injections](#)
[CVEs](#)
[Keys](#)
[Mails](#)
[Phones](#)

Show entries

Search:

#	Path	Date	# of lines	Action
14	/home/adulau/git/AIL-framework/PASTES/archive/pastebin.com_pro/2017/01/18/dgS8XHui.gz	2017/01/18	301	i q
30	/home/adulau/git/AIL-framework/PASTES/archive/pastebin.com_pro/2017/01/08/pYYL2bF4.gz	2017/01/08	22	i q
10	/home/adulau/git/AIL-framework/PASTES/archive/pastebin.com_pro/2016/12/31/hfxfRcW.gz	2016/12/31	20	i q
3	/home/adulau/git/AIL-framework/PASTES/alerts/paste.debian.net/2016/12/28/905218.gz	2016/12/28	463	i q
5	/home/adulau/git/AIL-framework/PASTES/archive/pastebin.com_pro/2016/12/26/ZQE3wTrg.gz	2016/12/26	243	i q
26	/home/adulau/git/AIL-framework/PASTES/archive/pastebin.ru/2016/12/23/hfxfRcW.gz	2016/12/23	442	i q
23	/home/adulau/git/AIL-framework/PASTES/archive/pastebin.com_pro/2016/12/19/YXIRuLv6.gz	2016/12/19	215	i q
28	/home/adulau/git/AIL-framework/PASTES/archive/gist.github.com/2016/12/14/openpanzer_1f83729b728b5e8f8e137e5f298514e1.gz	2016/12/14	0	i q

4024007171599474, CVV2, 107, 06/2018, Visa 4929070027209157, CVV2, 823, 01/2019, Visa 4916674750804533, CVV2, 422, 08/2020, Visa 377275329112778, CID, 4663, 11/2020, Visa 4716479018450414, CVV2, 549, 07/2020, Visa 4532896631347754, CV

AIL

Modules statistics

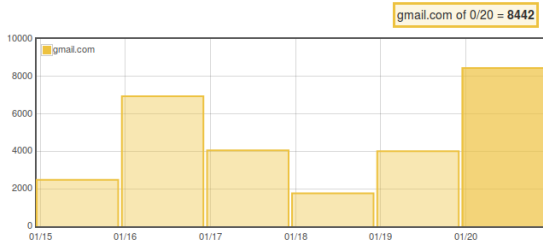
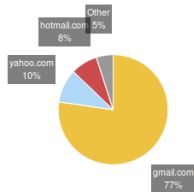
Browse important pastes

Sentiment Analysis

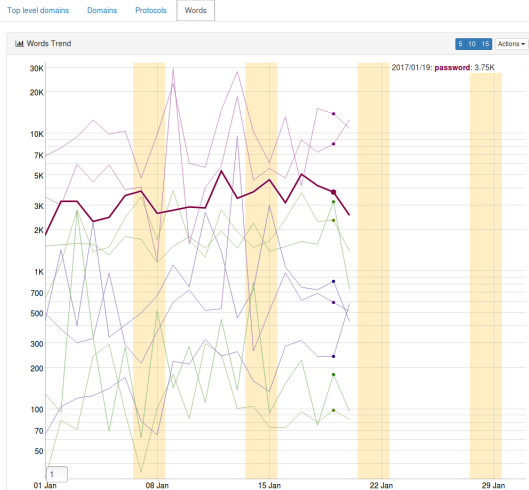
Terms frequency

Credential - most posted domain

Today



Trending charts



Terms frequency: Management interface

Manage tracked terms

Show entries

Search:

Term	Added date	Day occurrence	Week occurrence	Month occurrence	# tracked paste	Action
password	2017-01-20 14:30:51	2568	27064	85089	0	
visa	2017-01-20 14:32:06	578	10063	24709	0	
dump	2017-01-20 14:31:23	511	4165	21027	0	
mastercard	2017-01-20 14:32:19	70	6730	15697	0	
hacked	2017-01-20 14:31:39	171	3713	6658	0	
leak	2017-01-20 14:30:59	56	814	3923	0	





















Showing 1 to 6 of 6 entries

Previous **1** Next

Terms frequency: Top set information

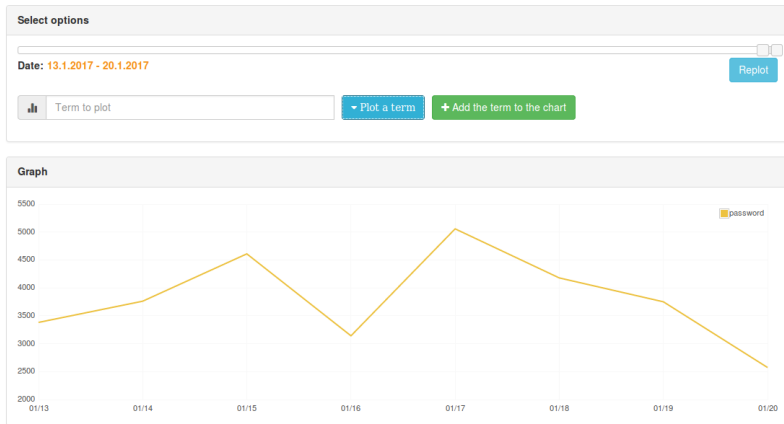
Today

Today top word

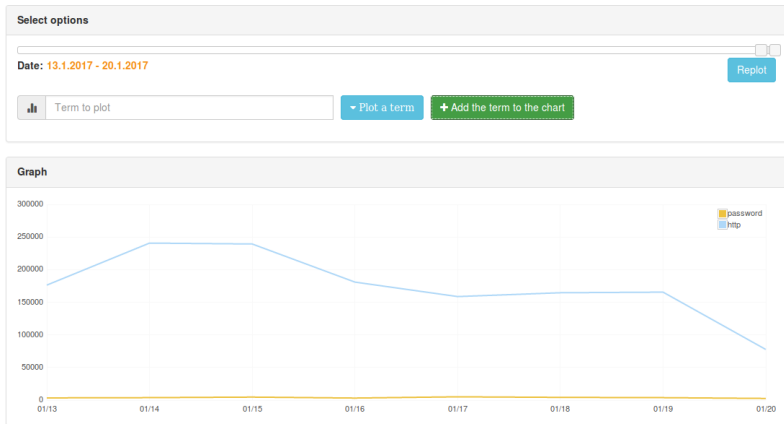
Term	Value	Action	Show	Position
http	77412	 	<input checked="" type="checkbox"/>	4, 4
system	56091	 	<input checked="" type="checkbox"/>	9, 8
hello	49575	 	<input checked="" type="checkbox"/>	<20, <20
extinf	44974	 	<input checked="" type="checkbox"/>	6, 6
string	37939	 	<input checked="" type="checkbox"/>	<20, <20
2017	36375	 	<input checked="" type="checkbox"/>	1, 1
filename	33703	 	<input checked="" type="checkbox"/>	5, 5
name	33281	 	<input checked="" type="checkbox"/>	20, 18
class	32503	 	<input type="checkbox"/>	<20, <20
live	31779	 	<input type="checkbox"/>	15, 11

Term	Value	Action	Show	Position
type	26057	 	<input type="checkbox"/>	<20, <20
return	25584	 	<input type="checkbox"/>	<20, <20
line	23080	 	<input type="checkbox"/>	10, 13
data	22833	 	<input type="checkbox"/>	<20, <20
index	22346	 	<input type="checkbox"/>	<20, <20
config	22051	 	<input type="checkbox"/>	<20, <20
rimworld	21623	 	<input type="checkbox"/>	<20, <20
from	21027	 	<input type="checkbox"/>	<20, <20
this	21006	 	<input type="checkbox"/>	<20, <20
true	20731	 	<input type="checkbox"/>	<20, <20

Terms plot tool



Terms plot tool

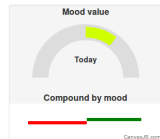
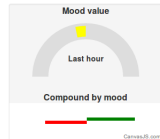
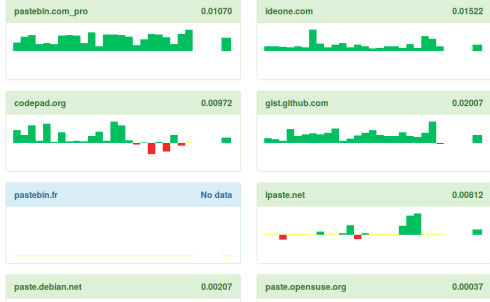


AIL - Sentiment Analysis

Sentiment analysis: Trending

⊕ Load data from all providers

Today's mood



AIL - Run your own instance

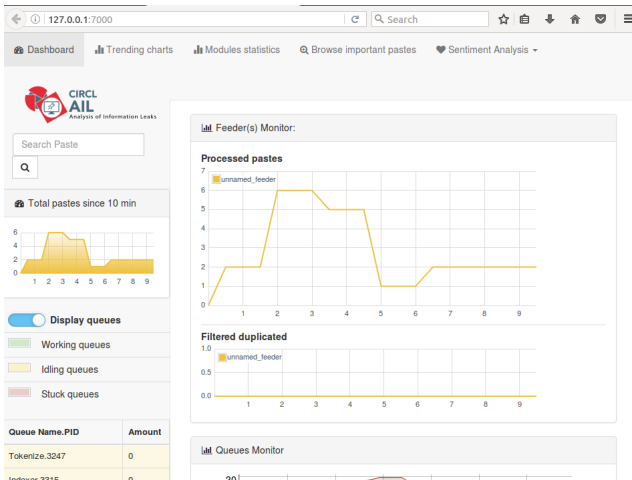
<https://github.com/CIRCL/AIL-framework>

The screenshot shows the CIRCL AIL web interface. The browser address bar displays '127.0.0.1:7000'. The navigation menu includes 'Dashboard', 'Trending charts', 'Modules statistics', 'Browse important pastes', and 'Sentiment Analysis'. The main content area features the CIRCL AIL logo and a search bar for pastes. A 'Total pastes since 10 min' chart shows a flat line at 0.0. A 'Display queues' toggle is turned on, with a legend for 'Working queues' (green), 'Idling queues' (yellow), and 'Stuck queues' (red). A table lists the 'Tokenize.3247' queue with an amount of 0. The 'Feeder(s) Monitor' section contains two empty line charts for 'Processed pastes' and 'Filtered duplicated'. The 'Queues Monitor' section is partially visible at the bottom.

Queue Name.PID	Amount
Tokenize.3247	0

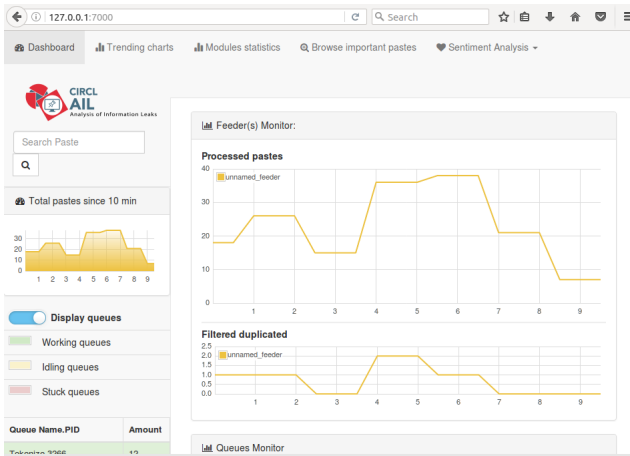
AIL - Run your own instance: With pystemon

<https://github.com/CIRCL/pystemon>



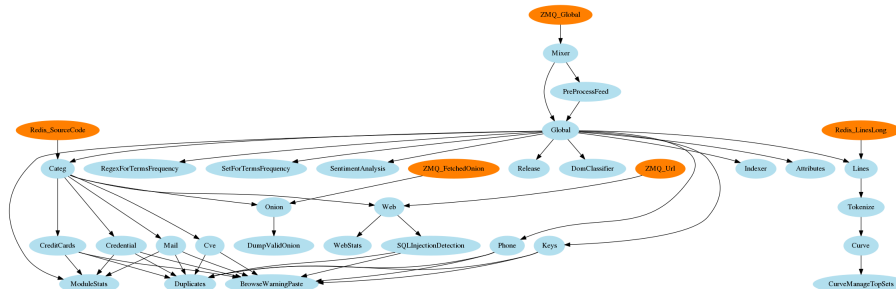
AIL - Run your own instance: Use CIRCL feed

Request access at: info@circl.lu



AIL - Add your own module

Choose where to locate your module in the data flow:



ALL - A sample module structure

```
import time
import re
from pubsublogger import publisher
from packages import Paste
from Helper import Process
if __name__ == '__main__':
    # Port of the redis instance used by pubsublogger
    publisher.port = 6380
    # Script is the default channel used for the modules.
    publisher.channel = 'Script'
    # Section name in bin/packages/modules.cfg
    config_section = 'Cve'
    # Setup the I/O queues
    p = Process(config_section)
    # Sent to the logging a description of the module
    publisher.info("Run CVE module")
    # Endless loop getting messages from the input queue
    while True:
        message = p.get_from_set()
        if message is None:
            publisher.debug("{} queue is empty, waiting".format(config_section))
            time.sleep(1)
            continue
        cveextract(message)
```

Conclusion

- Building AIL helped us to find additional leaks which cannot be found using manual analysis and **improve the time to detect duplicate/recycled leaks**.
- → Therefore quicker response time to assist and/or inform proactively affected constituents.
- Separating collection and analysis parts allowed us to extend the models in the CSIRT services.
- The modular architecture helped us to use the extracted data to **feed Passive DNS** or crawl Tor .onion.
- Ongoing work: Integrating AIL leak into MISP to **curate, share and collaborate on leaks**.

- Q&A
- <https://github.com/CIRCL/AIL-framework>
- Don't hesitate to contact us for feed access/exchange or ideas at <mailto:info@circl.lu>