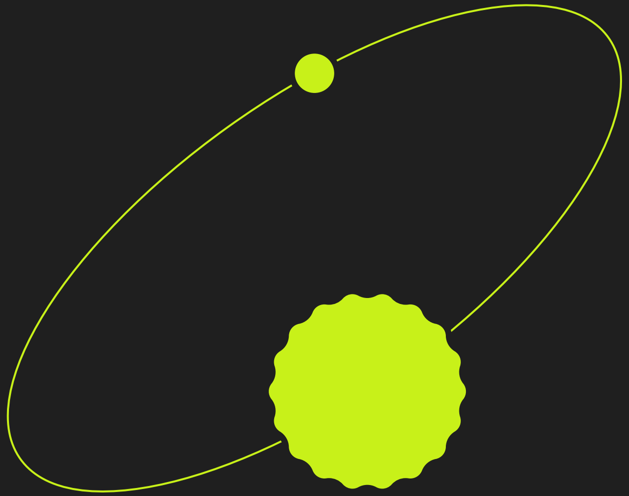
Cyber Threat Intelligence



A.

What is a data breach crawler?

Data crawling

"Data crawling is a method which involves data mining from different sources"

How it works?

Data Breach

Someone steals personal data.

Published

Someone buys the breach and it’s published.

Crawler

Crawlers store breaches.

What we use as example:

Have i been pwned?

Website that allows users to check whether their personal data has been compromised. It collects millions of data; the users can search by entering their phone number or email address.

Intelligence X

Search engine and data archive. It has unique features: the search works with selectors (email, URLs, IPs, etc.), it searches in places such as the darknet and others. Keeps a historical data archive of results.

Where did we find them?

1 **Clear web**

Using search engines and conventional sites.

2 **Deep web**

Deep web and onion pages, including black-hat sites.

3 **Undeground forums**

Undeground black hat hacking crime forums.

4 **Social sites**

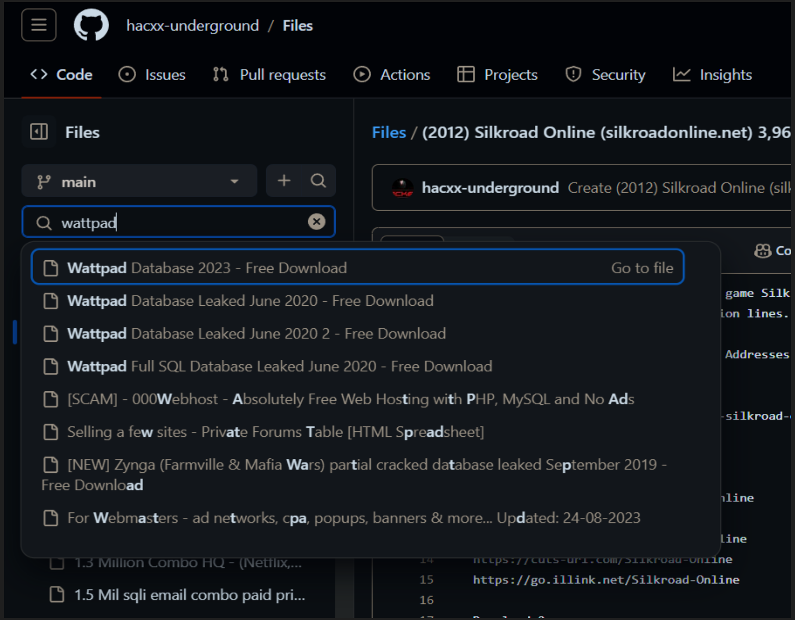
Social network and forums.

Breaches research

1 Clear web

By googling the names of the breached platforms, we can get download links.

They are often embedded in Pastebin pages, Github repositories for researchers, and even in clear web forums post.



Files / (2012) Silkroad Online (silkroadonline.net) 3,96

hacxx-underground Create (2012) Silkroad Online (silk

wattpad

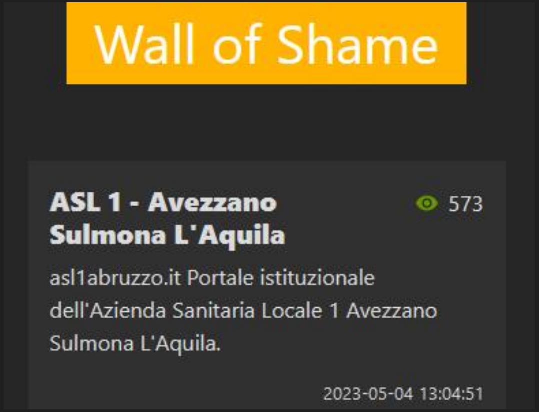
- Wattpad Database 2023 - Free Download
- Wattpad Database Leaked June 2020 - Free Download
- Wattpad Database Leaked June 2020 2 - Free Download
- Wattpad Full SQL Database Leaked June 2020 - Free Download
- [SCAM] - 000Webhost - Absolutely Free Web Hosting with PHP, MySQL and No Ads
- Selling a few sites - Private Forums Table [HTML Spreadsheet]
- [NEW] Zynga (Farmville & Mafia Wars) partial cracked database leaked September 2019 - Free Download
- For Webmasters - ad networks, cpa, popups, banners & more... Updated: 24-08-2023
- 1.3 Million Combo HQ - (Netflix,...
- 1.5 Mil sqli email combo paid pri...

```
game Silk
ion lines.
Addresses
-silkroad-
nline
line
https://cuts-url.com/Silkroad-Online
https://go.illink.net/Silkroad-Online
```

Breaches research

2 Deep web

On the onion sites of cybercriminal groups can often be found leaked information, often derived from victims who have not paid a ransom.

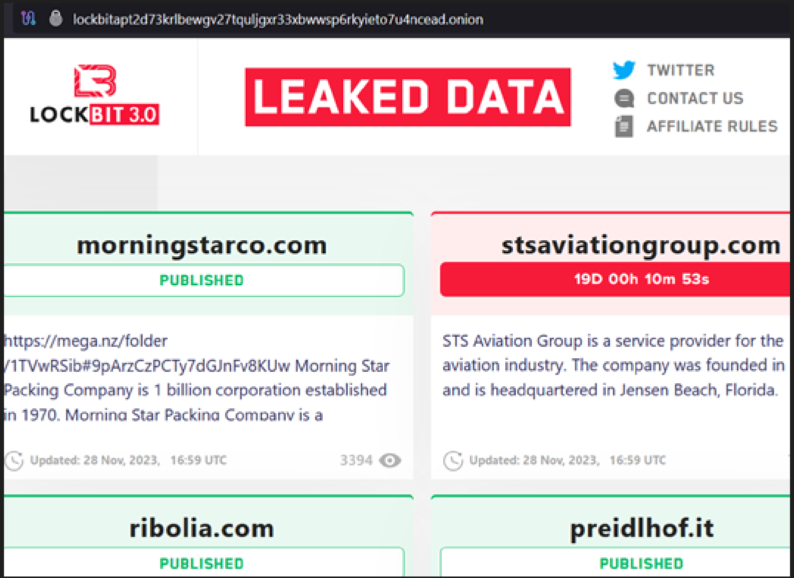


Wall of Shame

ASL 1 - Avezzano Sulmona L'Aquila 573

asl1abruzzo.it Portale istituzionale dell'Azienda Sanitaria Locale 1 Avezzano Sulmona L'Aquila.

2023-05-04 13:04:51



lockbitapt2d73krlbewgv27tquljgxr33xbwwsp6rkyieto7u4ncead.onion

LOCKBIT 3.0

LEAKED DATA

TWITTER
CONTACT US
AFFILIATE RULES

morningstarco.com
PUBLISHED

https://mega.nz/folder/1TVwRSib#9pArzCzPCTy7dGJnFv8KUw Morning Star Packing Company is 1 billion corporation established in 1970. Morning Star Packing Company is a

Updated: 28 Nov, 2023, 16:59 UTC 3394

stsaviationgroup.com
19D 00h 10m 53s

STS Aviation Group is a service provider for the aviation industry. The company was founded in and is headquartered in Jensen Beach, Florida.

Updated: 28 Nov, 2023, 16:59 UTC

ribolia.com
PUBLISHED

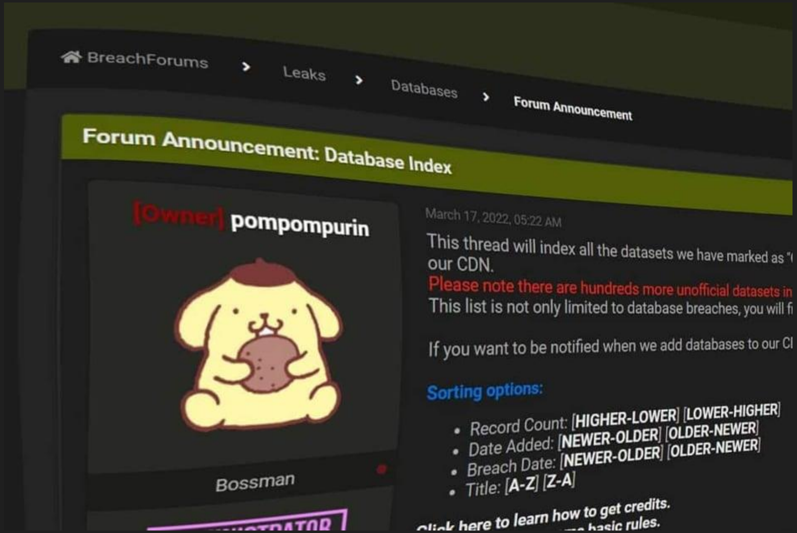
preidlhof.it
PUBLISHED

Breaches research

3 Underground forums

There are forums where cybercriminals publish breached data, often for free or with a system of tokens. Like Leakbase.io and Breachforums.

threat researchers without breachforums content to screenshot



Forum Announcement: Database Index

[Owner] pompompurin

March 17, 2022, 05:22 AM

This thread will index all the datasets we have marked as "
our CDN.
Please note there are hundreds more unofficial datasets in
This list is not only limited to database breaches, you will fi

If you want to be notified when we add databases to our C

Sorting options:

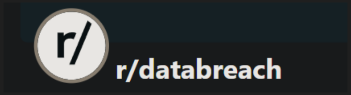
- Record Count: [HIGHER-LOWER] [LOWER-HIGHER]
- Date Added: [NEWER-OLDER] [OLDER-NEWER]
- Breach Date: [NEWER-OLDER] [OLDER-NEWER]
- Title: [A-Z] [Z-A]

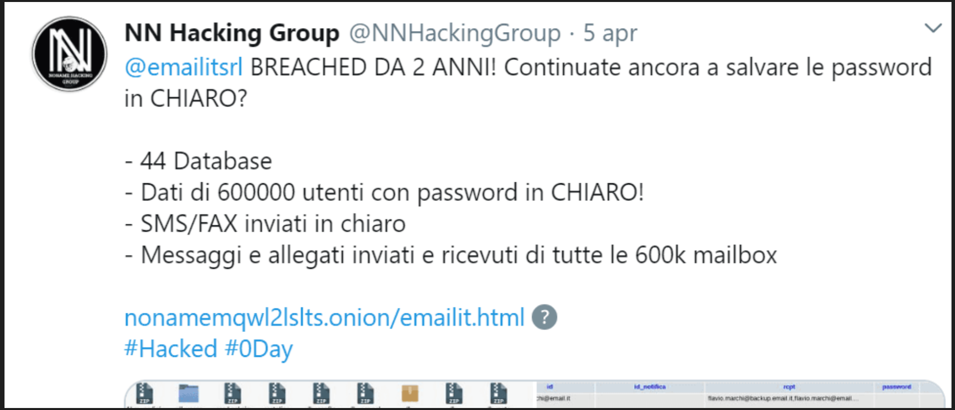
Click here to learn how to get credits.
basic rules.

Breaches research

4 Social networks and forums

Users often post links to leaked data download on places like Reddit, Telegram and Twitter.



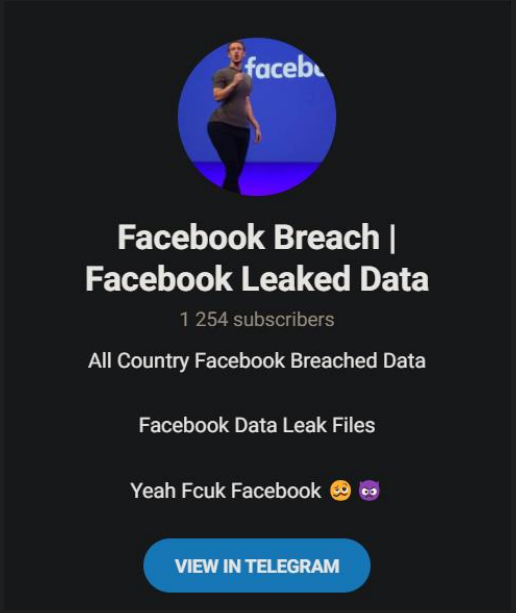

NN Hacking Group @NNHackingGroup · 5 apr

@emailitsrl BREACHED DA 2 ANNI! Continuate ancora a salvare le password in CHIARO?

- 44 Database
- Dati di 600000 utenti con password in CHIARO!
- SMS/FAX inviati in chiaro
- Messaggi e allegati inviati e ricevuti di tutte le 600k mailbox

nonamemqwl2lslts.onion/emailit.html

#Hacked #0Day


Facebook Breach | Facebook Leaked Data

1 254 subscribers

All Country Facebook Breached Data

Facebook Data Leak Files

Yeah Fcuk Facebook 🥺 😈

B.

The project

Data Breach Selection

Having limited resolutions, the group decided to use specific data breaches. The selection was made based on the information within them (telephone numbers, emails, passwords, etc.) and so that there were no redundancies between the various data breaches.

Data Breach Selection

Linkedin

ID, Email, password

Stockx

Email, name, surname, address, phone number

Twitter

Email, password

Edmodo

Email, password

Facebook

ID, name, surname, location, work, email, bithday, phone number

Functions:

Project functions

01 Password

Searched by API and our engine.

02 Search

Search for emails and phone numbers within files.

03 Identikit

Creation of the identikit of the chosen person using the information found.

Password search

We use the “Have I been pwned?” API for the password:

- The user enter the password.
- We create the HASH of the entered password.
- We take the first 5 characters of the HASH and “I have been pwend?” returns a range of HASH.
- On this return we select the correct HASH.
- The output will show how many times this pass is repeated in the data breaches.
- If the password is present in our databases, the table with the users using the entered password will be printed

Project Versions

Our project has changed for various reasons, here are the versions:

Version 1

HTML, CSS, JS

A group was already working in Python, so to differentiate ourselves we decided to use HTML, CSS and JS.

PROS	CONS
Server-less, DBless	Limited to 1 GB – 10mln record
Execute on Client	Depends on client resources
Easy to setup and customized	Non-scalable system

WE ACHIEVED 1 GB / 4 MLN RECORD

Version 2

NODEJS

To use when done so far, we opted to upgrade the system to NodeJS (Server)

PROS

Horizontal Scalability

Execution Speed I/O

NPM Modules

CONS

Memory Load Handling

Lack of Native Support for Relational Databases

Asynchronous Complexity Management

WE ACHIEVED 5 GB / 13 MLN RECORD

Version 3

PYTHON (FLASK + SQLITE)

Noting the difficulties in terms of resources and timing, we decided to base ourselves on flask and sqlite.

PROS

Rich Ecosystem and Libraries

Support for Relational Databases

Scalability

CONS

Performance in certain scenarios

Maintenance and optimization

Knowledge of different libraries

WE ACHIEVED +10 GB / +100 MLN RECORD

Version 4

APACHE SOLR

Apache Solr is the best performing solution for big amount of data, but it requires several hours of study and implementation

PROS

Docker

Multiple Hosts

Speed

CONS

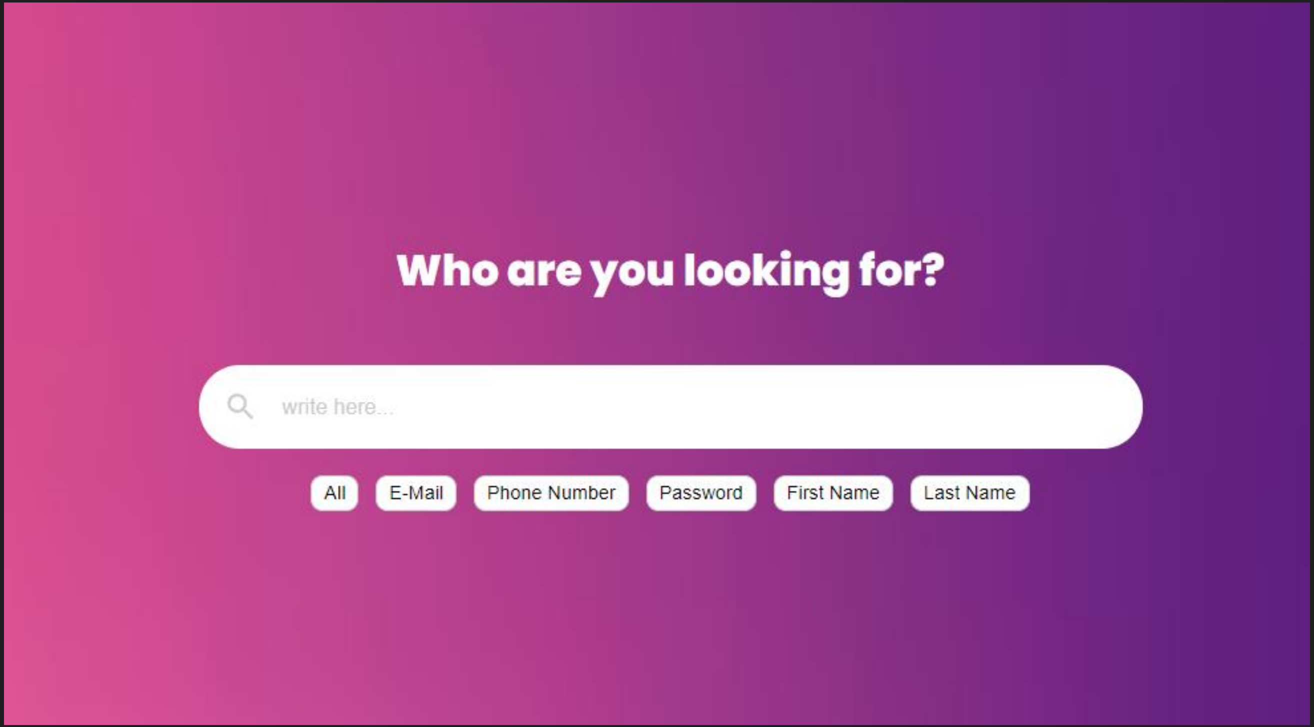
Difficult implementation

Ram usage

Data Processing

+ 13MLD

HARDWARE DEPENDENT



Who are you looking for?

The end

Any questions?