🗔 **ignis-sec** / **Pwdb-Public** ⟨ Public ⟩

A collection of all the data i could extract from 1 billion leaked credentials from internet.

⚖️ MIT license

☆ **3k** stars   ⑁ **393** forks   ⑂ Branches   🏷️ Tags 〰️ Activity

☆ Star	▾	🔔 Notifications

⟨⟩ **Code**   ⊙ Issues 5   ⑂ Pull requests   ▷ Actions   ⊞ Projects   🛡️ Security   ⠶ Insights

master ▾ | 1 Branch   0 Tags | | Go to file | Go to file | Code | ⋯

ignis-sec   Merge pull request #7 from 0xflotus/patch-1   ▪▪▪   4 years ago   ▪▪▪   ⟲

📁 statistical-lists	Added password|occurrence data	4 years ago
📁 wordlists	English is confusing sometimes. ...	4 years ago
📄 LICENSE	Create a license	4 years ago
📄 README.md	docs: fixed small error	4 years ago
📄 mystery-list.txt	Refilter mystery list to get more ...	4 years ago
📄 status.txt	Initial commit with 10k and 10M...	4 years ago

📖 **README**   ⚖️ **MIT license**   ☰

PWDB - New generation of Password Mass-Analysis

`repo size` `53.5 MB` `license` `MIT` `issues` `5 open` `issue resolution` `18 h` `open issues` `60%`

Goal

Leaving the 20 year old stuff of red team behind. Stuff works fine, and no one bothers to check/replace it through decades.

Included dumps

You can check the status.txt in this repository to keep track of included dumps.

Mystery List of 40k high entropy passwords

During my research, i've noticed a handful high entropy passwords (10 characters, uppercase-lowercase-digit) that were being reused. These passwords had really low occurrance rates, but it was still a lot more than i was expecting.

Some noticable stuff about these:

- They all start and end with uppercase characters

- They all start and end with uppercase characters
- None of them seem to have a keyboard pattern or meaningful word in them.
- They are all 10 characters long.
- They don't contain special characters.
- Some of them occurred up to 1 per 100 million credentials (meaning i have around 10 reuses of it currently)
- Most recent occurrence for these: 86 of these were found in a 55623 credentials from a leak in june 2020

I've filtered passwords which are 10 character long, and matches `(?=.*\d)(?=.*[a-z])(?=.*[A-Z])(?=^[A-Z][A-Za-z0-9]+[A-Z]$)(?!.*[a-z]{3})(?!.*[A-Z]{3})` which had an occurrence rating of less than 1.2 per 100 million.

~~I've released this list of 39576 passwords in mystery-list.txt under this repository.~~ I've refiltered it to get 763k passwords matching this pattern.

I have no idea what this uncovers and what it implies, but i'm suspecting a password manager out there is creating passwords with low entropy, causing repetitions over a lot of users. All the ideas about this are welcome and appreciated.

Curious about a statistic?

Please create an issue and explain what you want to learn, and if its interesting i'll query the thing and add the result!

Cool Stats

- From 1.000.000.000+ lines of dumps, `257.669.588` were filtered as either corrupt data(gibberish in improper format) or test accounts.
- 1 Billion credentials boil down to `168.919.919` passwords, and `393.386.953` usernames.
- Most common password is `123456`. It covers roughly `0.722%` of all the passwords. (Around 7 million times per billion)
- Most common 1000 passwords cover `6.607%` of all the passwords.
- With most common 1 million passwords, hit-rate is at `36.28%`, and with most common 10 million passwords hit rate is at `54.00%`.
- Average password length is `9.4822` characters.
- 12.04% of passwords contain special characters.
- 28.79% of passwords are letters only.
- 26.16% of passwords are lowercase only.
- 13.37% of passwords are numbers only.
- 34.41% of all passwords end with digits, but only 4.522% of all passwords start with digits.

Unique Passwords

- `8.83%` of the passwords are unique - they were only found once.
  - Their average length was `9.7965` characters.
  - Surprisingly, just a fraction of these passwords are meaningless.
  - Only `7.082%` of these passwords contain special characters - Rest matches `^[a-zA-Z0-9]$`
  - `20.02%` of these passwords are letters only, and `15.02%` is only lowercase.
    - Average length for lowercase-unique passwords were `9.3694` characters.

Language Specifics

I've partitioned my data depending on the top level domains of the email providers. (filter here: https://gist.github.com/FlameOfIgnis/9a1da894e8ae385a1ee58b8a734b8979)

I'm only releasing short lists of top 150 passwords for now. I'll eventually release full lists. I'll try to refrain from releasing an incomplete version publicly for now.

I'll update these lists per billion credentials i process. So even though some of the languages lists are not complete yet, they'll be in a better shape soon.

I've had enough data for

- Ukranian
- German
- Russian
- Italian
- Japanese
- Portugese
- Polish
- French

accounts to generate 1M password lists.

In contrast to that, I had too little data to work with for

- Serbian
- Azerbaijani
- Arabic
- Georgian
- Romanian
- Uzbek
- Macedonian
- Persian

- Armenian
- Albanian
- Turkmen
- Castillian
- Amharic
- Burmese
- Bulgarian
- Mangolian
- Catalan
- Tigrinya
- Tajik
- Korean accounts to generate reliable lists.

And i had a total of 0 accounts with Slovene language.

Cool Stats of comparison with rockyou.txt

- Rockyou.txt contains `14.344.391` passwords.
- After filtering most common `14.344.391` (same as rockyou) most common passwords, `11.583.476` of them were not in the rockyou.txt (a ratio of **%80**)
- Just in the first 1000 lines of rockyou and this data, there is a difference of **411** lines - meaning **411** passwords were not in rockyou.txt's top 1k lines.
  - From most common 1000 passwords, **37** of them were not in rockyou (all of rockyou). This passwords are:

37 passwords rockyou is missing that are ranked top 1000.

Very likely that around 8 of these are from test accounts/bad dumps that i failed to filter correctly

```
123hfjdk147
1464688081
159753qq
2012comeer
6V21wbgad
<password>
Blink123
D1lakiss
Exigent
Groupd2013
Indya123
N0=Acc3ss
R9lw4j8khX
Status
Telechargement
aobo2010
```

```
baili123com
bhf
cme2012
demon1q2w3e
demon1q2w3e4r
demon1q2w3e4r5t
exigent
g13916055158
hg0209
lincogo1
lizottes
megaparol12345
minecraft
nks230kjs82
nonmember
nyq28Giz1Z
pk3x7w9W
rr123456rr
startfinding
youbye123
yuantuo2012
```

Updating

I'll try to work in chunks of 1 billion credentials and update it regularly as it processes the data, until i run out of dumps.

Like the project? Help me throw more resources at it!

bc1quxwhewutde2ehjzqcflcgdjqwg34pmcdq3chcp

Releases

No releases published

Packages

No packages published

Contributors 2

ignis-sec Ata Hakçıl

0xflotus