Generating Personalized Wordlists With NLP by Analyzing Tweets

Utku Sen
Whoami

• Working @ Tear Security

• utkusen.com

• twitter.com/utkusen
Outline

- Small introduction to guessing attacks.
- Reducing the keyspace
- How to generate personalized password candidates
- Rhodiola tool / DEMO
Password Guessing Attacks

Offline

Online
Offline Password Guessing Attacks

- For cracking Hash, Encryption Key etc.
- More hardware resources = higher chance of success
- Only precaution is using slow algorithms
Online Password Guessing Attacks

- For finding website or application passwords
- More hardware resources $\neq$ higher chance of success
- Solid precautions

[Diagram showing the process of an attacker trying to guess passwords through an email server, with 'IP BLOCKED' indicating the successful prevention of an attack.]
We need to reduce the combination pool
Mask Attack

- Created by Hashcat team.
- “Try all combinations from a given keyspace just like in Brute-Force attack, but more specific.”

**Brute Force**
- **Charset**: mixalpha-numeric
- **Formula**: $9^{62}$
- **Combination Pool**: $13,537,086,546,263,552$

**Mask Attack**
- **Charset**: mixalpha-numeric
- **Formula**: $?u?!?l?!?l?!?d?d?d?d$
- **Combination Pool**: $237,627,520,000$
Still too much for online attacks.
The Sherlock Holmes Way
“Julia1984”

**Brute Force**
- **Charset**: mixalpha-numeric
- **Formula**: $9^{62}$
- **Combination Pool**: 13,537,086,546,263,552

**Mask Attack**
- **Charset**: mixalpha-numeric
- **Formula**: ?u?!?!?!?d?d?d?d
- **Combination Pool**: 237,627,520,000

**Sherlock**
- **Charset**: mixalpha-numeric
- **Formula**: ?
- **Combination Pool**: Usually less than 3
I suggested my favourite author George Orwell to my daughter Julia. I'm sure she will like it too.
Most of people are choosing words for their passwords based on personal topics*


Analyzing Myspace and Ashley Madison Wordlists

Analysis is conducted with PACK (https://github.com/iphelix/pack)
“a letter sequence is a meaningful English word if it’s listed in an English lexicon.”

https://wordnet.princeton.edu/
POS Tagging

• ~40% of Myspace and Ashley Madison wordlist are consists of meaningful words.

32% Singular Noun (NN)
3% Plural Noun (NNS)
1% Other POS-tags and Digit-only: Adverb(RB), Determiner(WDT), Adjective(JJ), Verb (VB) etc.
What We Have So Far?

- Leaked wordlists mostly consists of meaningful words.
- Passwords are based on personal topics.
What We Need?

Information About Target → ALGORITHM → PERSONALIZED WORDLIST
We need a data source
We need a data source

Cleaning Tweet Data

• Removing stop words (a, an, in, to, that, i, you etc.)

I suggested my favourite author George Orwell to my daughter Julia. I'm sure she will like it too.
Identifying Most Used Nouns and Proper Nouns

- POS tagging with Python’s NLTK library

I suggested my favourite author George Orwell to my daughter Julia. I'm sure she will like it too.

- NOUNS: author, daughter
- PROPER NOUNS: George Orwell, Julia
Pairing Similar Words

- Calculating semantic similarities of the extracted words.
- Wordnet’s path similarity algorithm is used.
- If the score > 0.12, we pair the words.
Finding Related Helper Words

I suggested my favourite author **George Orwell** to my daughter Julia. I'm sure she will like it too.

Rhodiola Tool

• Written in Python, mostly based on NLTK library.

• Takes a Twitter handle, generates a personalized wordlist

• Supported Languages:  English (Turkish and German soon)
utku$ python rhodiola.py --username elonmusk
...
tesla
car
boring
spacex
falcon
rocket
mars
flamethrower
coloradosprings
tesla1856
boringcompany2018
...
utku$ python rhodiola.py --username elonmusk --regex "(root|admin)\d{2}" --regex_place suffix
...
 teslaroot01
 teslaroot02
 teslaroot03
...
 spacexadmin01
 spacexadmin02
...
 tesla1856root99
...
boringcompany2018admin99
...
Mask Mode

```
TeSLa
CaR
BoRIng
SpACex
FaLCon
MaRS
EaRTH
...
```
Bring Your Own Data

```
uttu$ python rhodiola.py --filename mydata.txt
uttu$ python rhodiola.py --urlfile blogs.txt
```

carret-yetmedi mi
Elimize gecti bir kac kurus Para, Kasabadan Pazara kadar yapmadik zarar, toplandik, buyuduk, hepberaber ovunduk, goren olsa desin iste "bak su bir Turk!"
Ama suanda goryorum, gozlerime inanmeyorum, cunku bazı Türkler birbirini duşuruyorlar, fiyatlarla, silahlarla birbirine dusman olyurlar! Olur mu bizim amacımız yoksa bu mu? İnanmeyorum, hayir inanmak da istemiyorum, kötü Rüya gibi girdi bu benim kafama, bu üzüntülü yaklastiriyor beni mezartasma, sakin ha burakmayin beni bu skintyla yaltiz basima!
Ama beraber gücümüz yeter bu alcakliga, yabancı duyguya, yeni bir korkuya, yanyana beraber yeni yasanti kurmaya, Kardesce beraber yeniden yasamaya!

```
uttu$ cat mydata.txt

uttu$ cat blogs.txt

https://cobbr.io/Designing-Peer-To-Peer-C2.html
https://www.harmj0y.net/blog/redteaming/
https://danielmiessler.com/projects/ideas/
https://cobbr.io/Designing-Peer-To-Peer-C2.html
```
github.com/tearsecurity/rhodiola
DEMO

• A volunteer needed

• Requirement: Actively using Twitter in English
Conclusion

• Personal data can be used for password cracking

• Don’t use personal details in your passwords