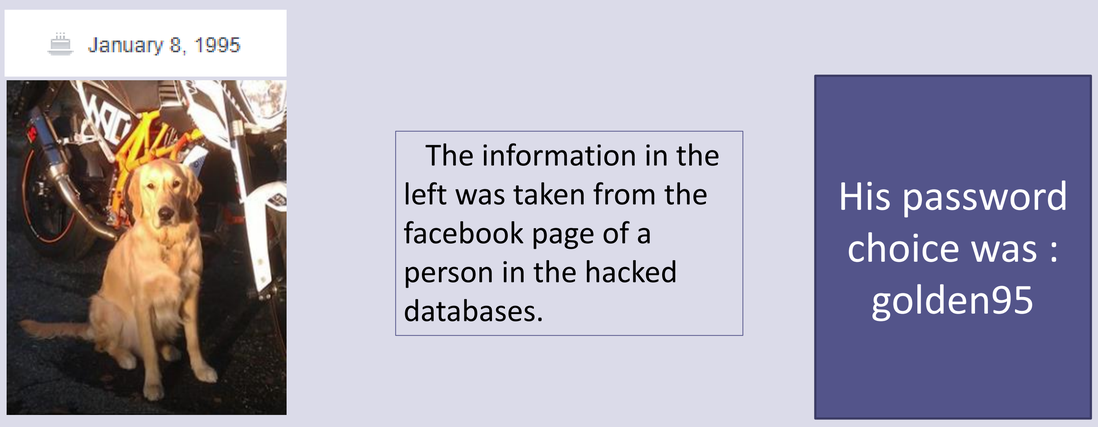
Personal Password Guesser

STUDENTS / UNIVERSITIES
Nazlı Mine Tülek

SUPERVISOR(S)
Albert Levi

ABSTRACT

The personal password guesser project's aim is to find out how many user expose their passwords or part of their passwords in the social media. Users generally share their birthdays, their relative's names, their favorite teams or their choice of music in their social media profiles. Since the users prefer to choose a password that they can easily remember, they generally choose a word that is personal to them, like the name of their pet , their birthday or a combination of these informations. These informations can be found in the social media profiles and if there is enough exposure of information in their profile, these can be used to generate a candidate password list for each individual.



OBJECTIVES

- Our objective was to see if the public appearance of a person in the internet affects his/her password selection.

INTRODUCTION

The project consisted of two parts. The first step was to create a collection of publicly available hacked databases. These databases were previously hacked and are available in the internet. In these databases there are email adresses,names or usernames and by using these informations we tried to identify the users social media profiles.

The second step was to take these social media profiles and extract keywords from them to create a candidate password list. After we created the candidate password list, we will try to match the password of the user to these candidate passwords.

DATABASE

This is an example of the personal information available in the database:

```
'city": "Milano",
'clearPassword": "lazzaro",
'email": "andrealazza@libero.it '
'firstname": "Andrea Lazza",
'username": "andrea lazza",
```

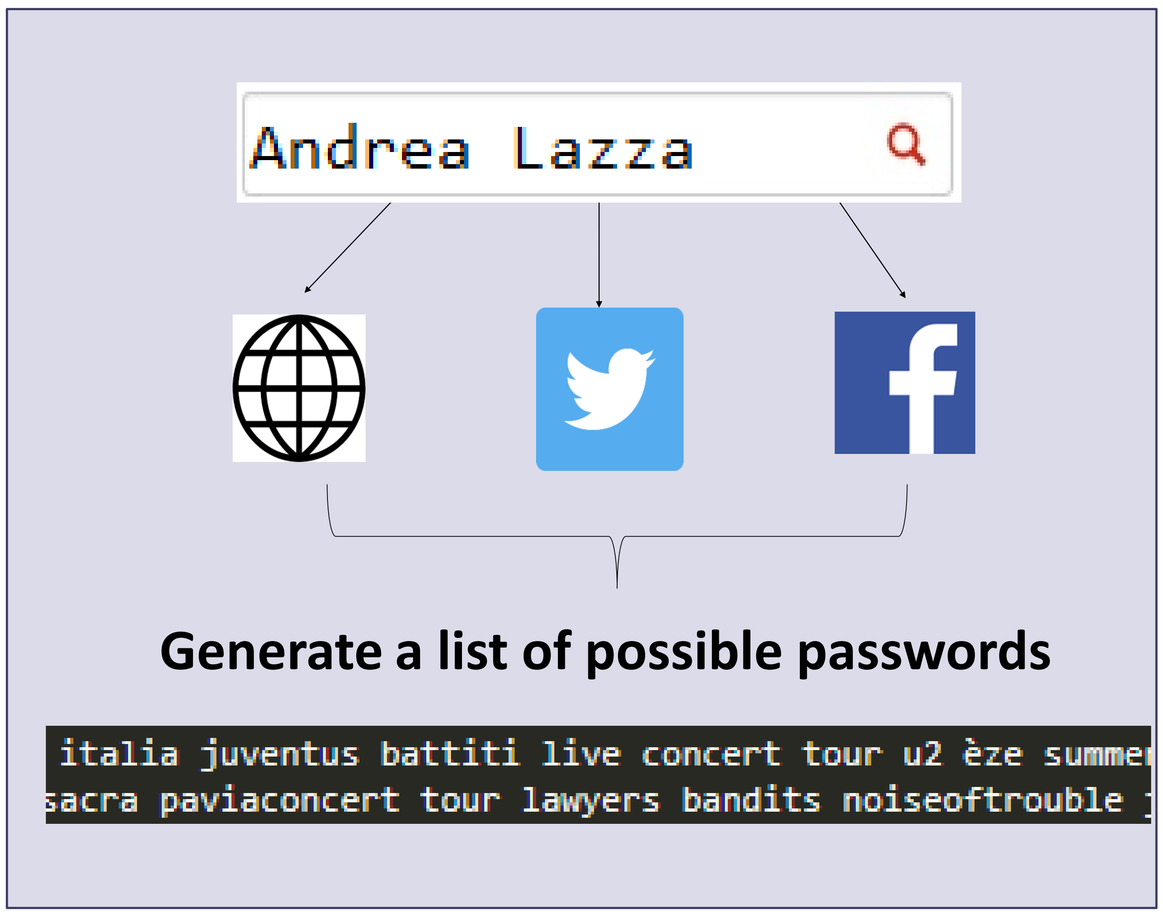
The first problem with the databases was to find enough information to identify a person in the social media. Most of the databases stored only the users email adresses and their passwords. But the problem is that, having the email adress of a person is not enough to identify a person. For example , given that the name of the user is 'Andrea Lazzo' the user may choose an email address like 'alazzo@...com' by merging their firstname's first character and their last name, making it difficult to identify the name of the user. Also even if they use their whole name, it is still difficult to identify them because there are generally multiple users having the exact same name. So we decided not to use the emails for identification because we need the name of the user and the emails are not an exact way to find them.

Instead we found other databases that store the firstname and the surname and also the cities that they lived in. The city information made itself very useful because even if we have the name of the user, we are still dealing with multiple people having the exact same names, and the city information made it possible for us to eliminate the false positives.

On top of that, another identification problem that we have encountered was , in some databases the people that we identified weren't active on the social media, therefore, even if we do find them, there weren't much information that we can work with so, we choose a database consisting of users that cared more about their social media presence. So we were much more likely to find personal information that they made public in their social media profiles.

Also, the social media platforms that we focused on identifying these people were Facebook and Twitter , because people tend to share more personal information on these social media platforms compared to other ones.

KEYWORDS



After we identified the users from Facebook or Twitter, we needed to retrieve their personal information from their profiles. For this purpose, we made a tool which enables us to pull data out of HTML files from Facebook and Twitter pages of the identified users in our database by using the BeautifulSoup library in Python.

In Twitter we scraped the tweets of the users which were easier to locate compared to Facebook because they are displayed on a certain page. But in Facebook the personal information is scattered through multiple pages. So we needed to automate the navigation around these pages in order to pull all the personal information that were available. Therefore, we needed to use another library named Selenium, which enabled us to navigate the necessary pages automatically.

The problem with web scraping was, Facebook disallows any scraping. But we are still able to scrape data from Facebook in small quantities. When we wanted to pull all the posts of the user from the Facebook or when we tried to pull data in a small amount of time, Facebook wouldn't allow it. So, we tried to to scrape the data in small quantities.

Finally after we retrieved all the data of the user, we merged the Twitter and Facebook data and we wrote a tool that takes the available hacked password of the user. The tool compares it with the keywords that we have found in the social media profiles.

CONCLUSIONS AND FUTURE WORK

Our main aim in the begginning stages of the project was to get close to the password by only using the keywords. We have tested 25 people and in this 25 people what we have found is the following :

- 3 exact matches
- 4 close matches

We have found that the 2 of the 3 exact matches used only their firstname as their password, and 1 of 3 used his Facebook username as a password.

In the close matches that we have found, 1 of them used a word extracted from his favorite quote and the 3 of them has generated a password with their firstname. 3 of the close matches had a number concatanated at the end of the password and 2 of these numbers were birth years.

We have tested 25 people so far but we will increase the identified users in the future steps of this project. Also, in the future we will improve this tool to get an exact match with the passwords by studying the password generation habits of users because most passwords are not formed by a single keyword .

REFERENCES

Tam, Leona & Glassman, Myron & Vandenwauver, Mark. (2010). The psychology of password management: A tradeoff between security and convenience. Behaviour & IT. 29. 233-244. 10.1080/01449290903121386.