**Abstract - The aim of this project is to create a password universe (PU) modelled on a web application. This PU is used to visualise a given password database as the “universe” and the entries of it as “stars”. The user is treated as the centre of this universe by choosing a keyword to compare with the database. Database doesn’t necessarily need to be with passwords, it can be any combinations of strings such as an English dictionary. It’s mostly aimed at password researchers and cyber security experts allowing them to see patterns thus helping them to gain further understanding of how passwords are related to one another. It can allow users to compare the same password with different databases and gain results of how users come up with their passwords.**

## **Introduction**

There are billions of accounts hacked every year, unhashed data being spread around the internet that can be gathered if you know where to look. For example, recently there was a leak of millions of emails and passwords stolen and a lot have been unhashed and are spreading around [1], making hacking users easier using standard password cracking techniques such as dictionary attack, brute force attack and rainbow table attack. However, there is also a problem that with such huge database of real unhashed user passwords, that many other can be potentially at risk because they have had a similar or the same password as the one that got hacked causing a leak of for example 10 million, possibly affect a much larger population. Alternatively, even if the user is aware that his password has been hacked, he will use a simple pattern to change it. This project aims at using any password database to visualize potential patterns people have whilst creating their passwords which can be used to potentially strengthen the requirements for passwords on websites. For example, research such as this can be used to improve something most websites have, password meters. How does one decide what rules to add to their password requirements for these password meters? Different password meters perform differently [2] [3]. There was a research made in 2010 stating that our current password creation policies are still vulnerable to online attacks [4] and this issue still persists as more and more users are keeping the same passwords [3], using the same patterns of creation without much knowledge of them being hacked and vulnerable, and as I mentioned even if they know, most will result to a simple pattern fix.

When starting this project, I had multiple scenarios presented to me for data visualisation that I can approach such as;

1. Heat maps – generalised heat map giving the frequency of each character at each position and each password is a segmented line on the heat map.

2. Human related passwords – passwords related to geo-locations, related to a persons name, animals, plants and even religion, languages and wars.

These two are just some of the many ideas I was given by my supervisor but I chose to go along with the “Passwords in a universe” where each star is a password and the user keyword is in the middle of this universe.

The reason behind me choosing this idea is because personally I could already see the possibilities of how much this can benefit professionals, I could visualise what I want to build and how I would like to make it work, the issue was how I would achieve that. Nonetheless I believe I made the correct choice and I hope this tool can help improve and break patterns that current users have with regards to their passwords.

I believe that with the power of visualisation, it can be much easier to actually see patterns and collect data otherwise left unseen. The use of visualisation techniques have been used for a very long time from maps to visualise one’s location to graphs, to visualise one’s data. In a short Q&A video, Simon Samuel, Head of Customer Value Modeling for a large bank in the UK, he answers the question “How important is data visualisation” by saying:

“Visualisation is fundamental and it will be increased with its usage going forward. The executive of the bank where I work currently demands more visualisation tools to help them support their insight and analysis and also to accelerate their understanding”.

Visualisation can identify areas needing for additional attention much easier and simpler for the human brain to see and identify. Any visualisation tools that can help understand data and make important decisions on how to detect patterns and try to fix them can be useful

(<https://dl-acm-org.chain.kent.ac.uk/citation.cfm?id=2525496>)

In this paper I will be talking about:

1. The issue, how users don’t understand the risks of using same passwords everywhere, how even if some of their passwords are slightly different, they use simple well known patterns to differentiate them which can easily be spotted using my visualiser.

2. My password universe, what can my software do, how can it potentially help and show some examples of known patterns, how does using edit distance improve on today’s password meters algorithms which look at stuff such as “symbols, capitals, variety of letters” etc.

3. Results, I ran some benchmarks to look at speeds varying on different database sizes. How the size of the keyword affects the speed.

4. issues, what are some of the currently known issues and how they can be improved.

5. Conclusion, what I learned in the end, how helpful this can be.

**2. User patterns, no true password entropy**

Simple text passwords are the most used form of identification and grant permissions. Even though there are many other ways of identifying yourself such as face recognitions, thumb prints, key cards etc. most places still use the standard password. By default if something is used the most, it should be the most secure in terms of identification but in regards to password cracking, its currently a big issue because of how users create their passwords and once leaked somehow, even if warned, they persist to make simple changes instead of making a serious change in their passwords.

### 2.1 User Patterns

Most users when creating a password have fallen into some known patterns that they use assuring themselves that their passwords are secure. An issue comes from the overly rated password meters which allure users to believe that since it satisfies the meter, their password is surely secure and uncrackable.

(<https://www.researchgate.net/publication/276113338_Cracking_More_Password_Hashes_With_Patterns>)

(<https://wpengine.com/unmasked/>)

For all examples below in this section, I will be using a standard initial password : “password”.

There are a few common patterns that people use to make their password more secure. For example, capitalising the first or last letters of their password. One of the most common password meter requirements is to have one upper and one lowercase letter. Users are so used to their current passwords that they choose to adapt their current password by just capitalising a letter in their password that would be easy for them to remember but yet satisfy the password meter. For example, turning “password” into “Password”, “passworD” or “PassworD”.

Furthermore, there is also the “1” or any short combination of following numbers at the end of the password. Again this comes from the standard requirement of password meters requiring at least one number into your passwords giving false sense of security to users thinking their password is securely protected.

For example, changing the initial password to “password1”, “password12” or “password123321”

Finally, there is another common pattern people use as an alternate or with the one stated above which is replacing letters with similar looking numbers such as (t → 7, s→ 5, I → 1, o→ 0). For example, “password” would turn into “pa55w0rd” or “pas5w0rd”.

Combining all of these tricks the standard “password” would turn into potentially something like “Pas5w0rd123” which from the normal users perspective makes him believe that this change of his password. Password meters are evolving but they are doing small changes which just pushes users into more patterns which eventually makes no difference of their password being cracked.

These are some of the most known patterns and if one of a users old passwords has ever been compromised by a leak or any other way, there is a high chance the person trying to crack you will try to follow all of these pattern combinations. These examples show how users are overly relying on just satisfying a password requirement leading to ideal security for their passwords. This calls for a drastic improvement on password meters, something has to change and the ideal way is to notice as many patterns as we can that users are creating and proposing solutions to these patterns so we can tighten securities.

### 2.2 Adapting password meters

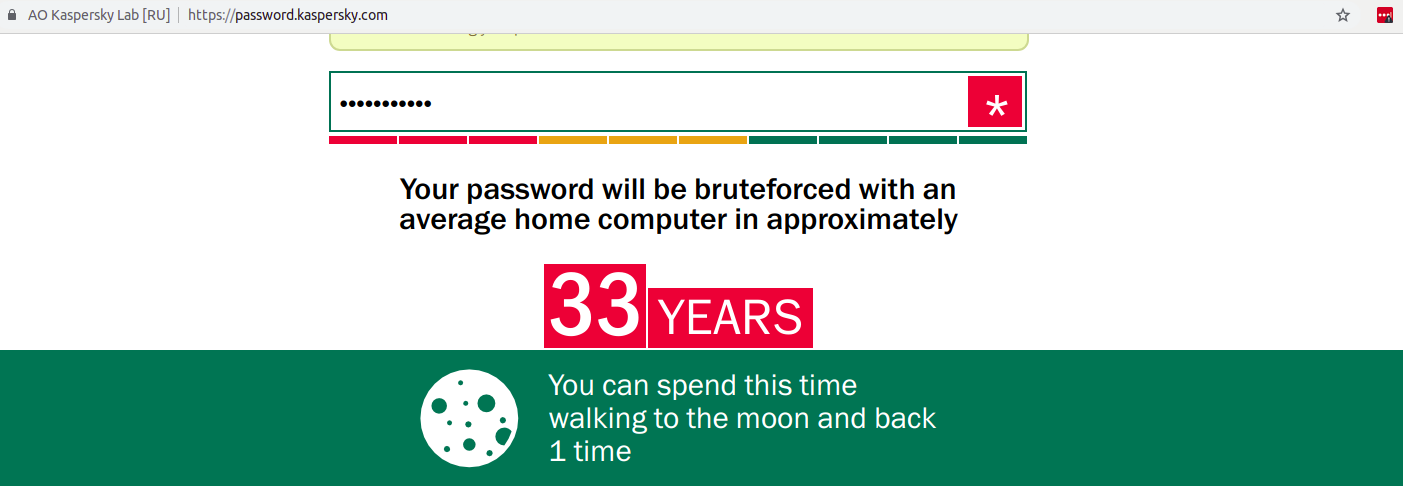
Password meters have constantly tried to evolve and add more criteria into them or changing the algorithms to portray how strong the password actually is (<https://dl-acm-org.chain.kent.ac.uk/citation.cfm?id=2989058> ) but even with all these changes, users just start creating more patterns which are not addressed which causes users to see what the strength meter says on the password meters thinking its highly secure password but in reality, its a small change.

For example:

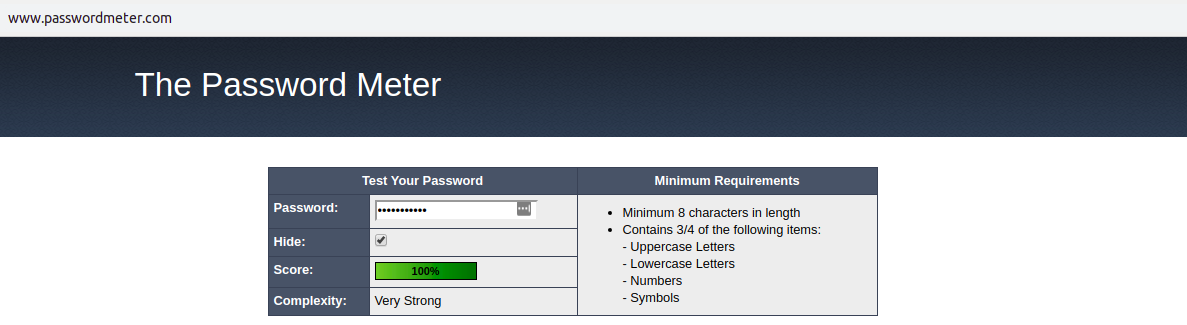
Let’s take a short combination of my DoB and name and satisfy a password meter. I will be using “9970707gG!1” Where “997” comes from the year I was born 1997, the “0707” are day and month, “gG” comes from my first name letter, and just capitalising it to make the password requirements happy, and “1!” which is just the first number on my keyboard to again just satisfy the password checker. Now this might seem secure but in reality, this is just a mix of simple patterns to update my original password “9970707” which is what I had at the beginning and its fairly easy to guess if the person trying to crack me has specifically targetted me and can figure out my DoB quite easily from social media.

Let’s see what some password strength checkers say about this password.

1) <https://password.kaspersky.com/>

2) <https://howsecureismypassword.net/>



3) [www.passwordmeter.com](http://www.passwordmeter.com/)

These password strength meters were taken from the top results google offers when you search for “strength password checker” and which is what most users would probably end up using if they want to check the strength of their password. And as complex as this example password might look “9970707gG!1”, if my old password “9970707” was leaked, someone who has knowledge of these simple patterns can easily crack this password and gain access showing how misleading password meters can be.

​​// maybe create a section talking about different password meter algorithms, how they determine if a password is secure.

## **3. The Password Universe**

Unlike password checkers which use some algorithm to detect the total amount of symbols, numbers or other variables within the password to determine its strength, I use an edit distance using Levenstein distance formula to determine the total number of mutations a password needs to take to go from A to B or in this case, from the keyword a user typed to every database entry. All databases that I have used are online and are unhashed plain text leaked passwords that were made available for research purposes.

### 3.1 Levenstain Distance

The benefits of using levenstain distnace is that it completely bypasses all the tricks users use as patterns and it’s much easier to see what they are doing. How it works is that it looks at all changes that need be done to get from one password to another so if your password is “password” and you change to for example “[p@ssword](mailto:p@ssword)” or “p4ssword” since both “@” and “4” are known replacements for “a”, a password meter would determine them in different strength since it would believe that having @ instead of 4 or a would be better, alternatively if you already have a symbol but no number, it would think having 4 better than @ but in reality, its a single character change to your password.