Large Scale Research on Password Practices

Joaquin Valdez
82370040

ABSTRACT
This project aims to analyze commonalities between users passwords . For this , I will take a sample of about 40,000 passwords that protect credit card information. The sample will be tested against all good practices of giving passwords.

INTRODUCTION
For years, people have unsuccessfully tried to figure out the best way to secure personal information. However, because of the nature of the WEB, passwords are the only the only reliable, cheap and accessible way to secure information on the WEB. Personal passwords are meant to give access to the owner of the account, hence creating a barrier for all other users.

Hackers have tried to figure out complex ways to crack passwords with mechanisms such as; brute forcers and dictionary attacks. To prevent this,, today most websites disallow multiple log in attempts. In response to this , crackers then, have began or tried to hack the website itself or the user in other to obtain the password.

However, the average user may be able to get into an account without using a computer worms, SQL injections, a key loggers, etc. They may figure out a password just by an educated guess or by having access to any type of personal information. One may expect that in the year 1998, web passwords would have been rather weak, since people were naïve and beginners with the world wide web technology. It would be logic to infer, that years later, and with more expertise on the subject, passwords would be stronger. This is what this project will find out. Also, This project will realistically provide an insight into why people fails to provide strong passwords.

ABOUT THE PROJECT
I have work as a J2EE junior developer for eight months now for a medium sized website with an average daily traffic of about 6 to 8 thousand people.

The site counts with more that 300,000 customers from north America and Europe. The costumers must create an account with the site, requiring personal and credit card information prior registration. During my time working there, I have realized how careless people are when providing passwords. Hopefully, this study will help people realize weaknesses in their passwords.

The site counts with about 100 independent sellers and other management staff (non IT people) that “in theory” should have secure administrative passwords.

For testing proposes, I count with direct access to less than one sixth to the total amount of accounts of the website. However, because of the nature of the information and confidentially reasons, none of these will be presented of the report, nor during the presentation.

PREPARATION
All the material is in my personal SQL server. For security proposes, I have encrypted them. The list of 43,777 accounts which include user names, passwords and email addresses for actual users recorded over the last four years.

Since analyzing such huge amount of passwords is a difficult task for the average human, I have coded a software that will parse and analyze passwords against many different variables such nicknames and emails that make them guessable.

THE LENGTH OF PASSWORDS
First, let’s differentiate between the basic scheme for classifying weak and strong passwords. Basically, the more characters la password consists of the better protection it offers. However, this rule may not strongly apply to websites. Since protection efficiency is based upon different combination, not solely on length.

The production of weak passwords is recurrent, given that websites do not enforce the password length as a requirement, passwords can be typed with any length.
The results show that close to 164 people used 2 letters or less as password to protect their credit card information. 278 people used 3 letters and 1688 used 4 letters. Amazingly, 4.82% of the people either are not aware that short passwords do not offer enough security or do not care that their credit card information is vulnerable.

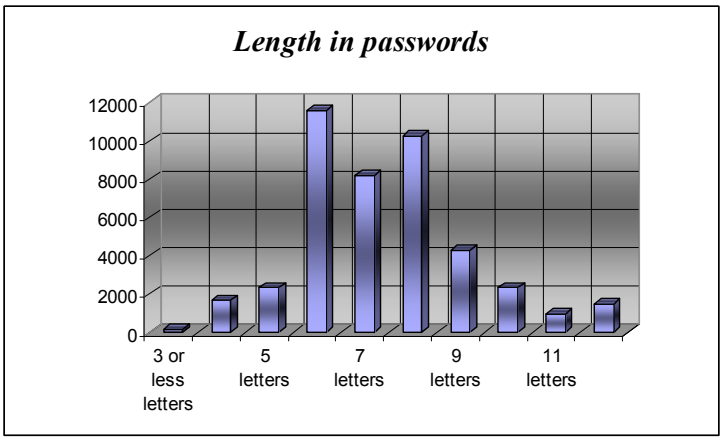

Figure 1 shows the frequency in the length of the passwords

However, the good news is that apparently people are learning more and more on how to provide more secure passwords. The sample shows that two years ago 5.4% of the people had a password with at most 4 character; today, this number has gone down to 4.15% .

THE UNIQNESS OF A PASSWORD
In the world of phishing, it is well known that there are some password that are much more common than others. People still choose passwords that do not offer security and follow a predictable patron.

One can not believe that the sample throws illustrates that up to 200 people chooses the word “password”, 126 people choose the famous “123456” number combination and that 36

of the users choose “1234” as their password. This means that almost 1 of every 150 people or 0.8 % uses either the password “password”, “1234” or “123456”.

Also, interesting is the fact that people tend to use clues from the website in order to select a password. This may help them remember the password making it easier of them, but it’s a backfiring effect because they became an easy target. For instance, data shows that 125 people uses “tickets” as their password while 56 use “hockey”, 27 use “baseball” and 30 use “soccer”. Because selling hockey, baseball and soccer tickets is a very important feature of the website, one may suppose that in fact these two can be used as a password. Other interesting fact is that the slogan of the website is used 60 times as password.

So far between these 6 easily guessable passwords, we have almost 2 % of the total sample. However there is more; an other easily guessable password is the name of the website. The name of the website is a common but weak password selection strategy of many people. The sample throws an impressible 200 people that uses the name of the website or some small variant of it as password to add up to 2.2% of people with easily guessable passwords.

The password “qwerty” seems a strong random password hard to guess. However, 27 people has this as their password. I had no clue why this was such a common password until I looked at the keyboard. In the keyboard, the letters Q W E R T Y are continue in a row. This clearly shows the complete lack of care of some people when selecting a password.

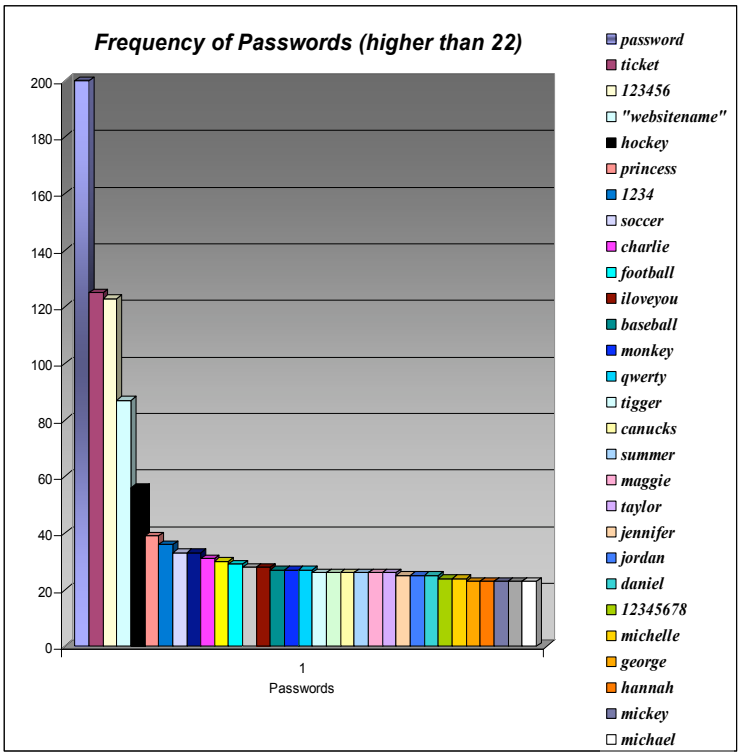

Fig. 2 shows the frequency of the most used passwords in the sample

Seen the previous results , it came to no shock to discover that people proper name counts for a huge quantity of passwords. Either the account owner’s name or the name of someone related to him/her. Daniel, George, Michael, Charlie, Mickey, Taylor and Jennifer are some of the most common names used in the English language and not surprisingly these are some of the most common passwords used. Each has an average usage of more than 24 repetitions. All of them make up to 297 people who use these name as passwords.

Finally, between all these easily guessable passwords based on communalities of all passwords, there are 1172 in a sample of almost 40,000 accounts for a 2.9 % or 3 of every 100. Other weak common passwords found but a little harder to guess are nouns such as money, tiger, etc….

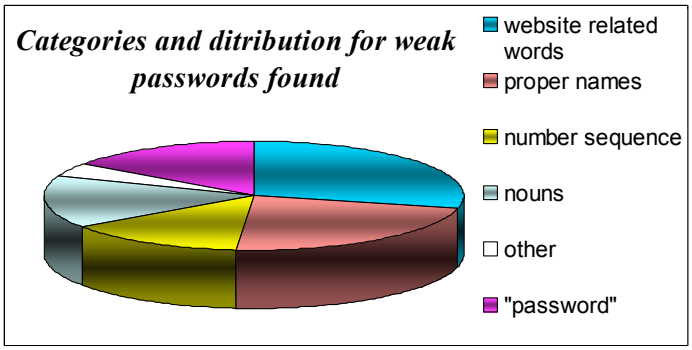

Fig. 3 shows the distribution of weak password given by category

NICK NAME RELATED PASSWORDS

Humans easily forget passwords so establishing password that is related to their nickname is highly functional, but offers very limited protection.

Certainly, choosing the same nickname and password is foolish when it comes to security; however, in the sample, 1174 users they did just that. If it makes no sense to register for this website unless, they are going to make a purchase and therefore, requiring to provide your credit card information, hence, protecting credit card information with a functional , but extremely weak password; you may as well hand in your wallet to the first person you see on the street.

THE BRAIN AND ITS INCAPACITY TO REMEMBER DIFFERENT PASSWORDS

In a paper from Dinei Florencio and Cormac Herley from the Microsoft research center, they find that

“Each user has about 25 accounts that require passwords, and types an average of 8 passwords per day.”

Because the short term memory for random passwords is extremely limited in our brain. People tend to use the same password for different websites. Non IT people do not know that on every website, the staff has access to their account information.

Furthermore, hackers have been attempting to hack small websites and steal user’s accounts. The stolen accounts won’t be used to access the website but rather to try to match the password and the email provided in hope the user uses the same password.

To exemplify the dangers of using the same password. I took a random sample of 50 user accounts. I took their passwords and email addresses provided, and attempted to access their email yahoo or hotmail accounts. To my surprise, I was able to access exactly 32 email accounts. For privacy reason, I did not access any email from these people. However, 25 of these accounts seemed to contain important information such as ebay, paypal and other personal emails while the other 7 email accounts seemed to be either abandoned or full of meaningless spam.

When I started this research, I was skeptical about finding significant weaknesses password because of the great deal of information that flows around the web of how to choose a secure password. However, the more I continue with this research, the more impressed I am on the vast weak passwords out there.

DICTIONARY PASSWORDS

To test how many people now days used passwords commonly found in the English dictionary. Since, writing a program to match 40,000 passwords to words found in a dictionary would be very inefficient. I used my SQL database with more than 1,000,000 words commonly used on dictionaries. The query: "select distinct i.uid, o.uid from tuser i, tdictionary o where o.word = i.password" produced the result in only a few seconds.

The result were surprising. 12620 passwords would be found in common dictionary so this leaving 1 out of 4 passwords coming directly from a dictionary. This is not a big security treat for a website since multiple log in tries can easily be disallowed; however, we can expect people to use the same patron when assigning passwords to systems that may in fact be vulnerable to dictionary attacks.

ADMIN PASSWORDS

With a limit sample of about 150 admin and staff passwords, the results obtained follow.

All passwords used by IT people were random and with a length of at least 7 characters. This is expected from people who know the dangers of weak passwords; however, other members of the staff fell into the categories given bellow. Giving about 4% of them with weak passwords and also, 1 of every 4 uses passwords commonly found on dictionaries.

SOLUTION

Although, this report showed that people is learning to use strong passwords, the rate in which they are learning is very low (1% of people uses stronger passwords two year latter). The best solution is that websites enforce scanning passwords protocols. First, they must enforce user to provide passwords with more than 4 character length. Second, passwords must be scanned against other passwords to find, those that have been constantly used and disallowing them and the website must warn the user to select a unique password, for this specific website. Also, encrypt all passwords before storing then on database incase accounts are leaked. Websites such as Paypal, and Ebay that contain very confidential information, must prop the user a set of personal questions when he/she registers for an account. Then, every time the user provides the correct password a randomly selected personal question (from the set answered before) must be responded before been able access the account. This would ensure that if for some reason the user uses the same password in an other website, whoever tries to use the password to force his way in into the user ebay, or paypal account fails because the answer to the personal questions provided by the user at registration wont be known by the attacker.

CONCLUSION

Amazingly, after more than 10 years that the internet has been essential in the daily life, many people still do not know how to protect their personal information from others. After looking at this research, I believe that websites must educate people in how to select strong passwords to keep their information safe and away from others. This also benefit the website because yearly, there are many complains from used that claim that their account was broken into. Lengthily, composed by random letters and numbers, and unique, passwords must be in order to be strong and safe.

BIBLIOGRAPHY

http://www2007.org/papers/paper620.pdf
Dinei Florencio and Cormac Herley, **A LargeScale Study of Web Password Habits.**

There is no other references since this project was an analyzes of a test sample

The source of the sample is can not be publicly know. However, the professor will validate it.