Is your username revealing your password?

Gantela Prabhakar and Cheldi Sharath

Department of Computer Science and Engineering,
SreeKavitha Engineering College (SKEC),Karepalli, Khammam (AP), India

Abstract- In general when security is concerned with your personal data or any other details that you want to be secured; we take outmost care in hiding the password so that the data is protected from unauthorized users such as hackers, intruders etc., But the prime job in stealing the password of a user is to understand the encryption method followed, the algorithm implemented in it or the key that has been provided for it. In this paper we discuss another angle of revealing the passwords which follows a little bit common sense added with some letter frequency in guessing the passwords which doesn't need any key, the scheme of implementation and also the encryption method that is used.

Index Terms- dictionary attacks, letter frequency, password management, relative frequency

I. Introduction

The classic view of authentication from common mans angle is to secure his data with a username which may be public and a password which is purely private. Most of the transactions that are being carried out now-a-days are through online transactions where for each interface a login page is displayed. The user has to register him self so that he is identified as an authorized user to that database or the website in order to access the private data of that organization

With respect to this the registration process is done for every online transaction that now-a-days an easy way of reducing the human risk. E-mail accounts, online banking, e-business and few other operations that are carried out now-a-days are so casually done that they are at few finger tips away from the user.

In this paper we try to concentrate a method of avoiding the attacks called the dictionary attacks which are prominently improving its impact. Even though we have many mechanisms to avoid unauthorized access to personal data, the dictionary attacks may not need any firewalls or any other cryptographic mechanisms from accessing our personal data.

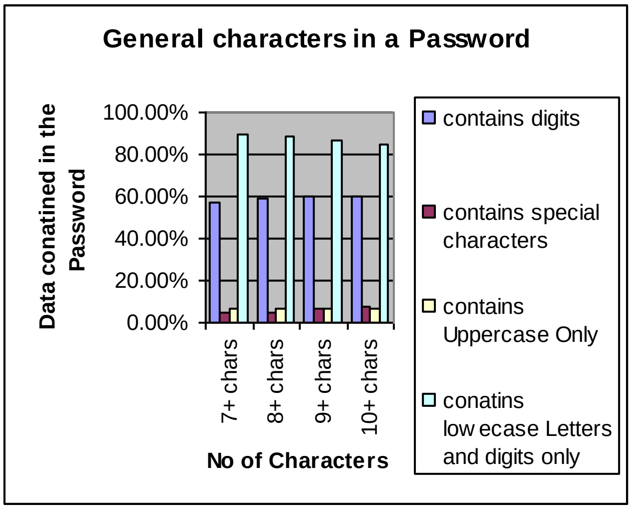
II. General Methodology

We begin our discussion with general process of registering into any internet based transactions website. The user needs to register him in order to perform online transactions. When the user is about to register the website asks him to fill the details in the form displayed on the screen. The form contains few fields which may include his name, e-mail id, mobile number, address, preferred username and a password of his choice. On filling the required fields he will be now able to perform online transactions. Now when the mandatory fields are entered in a proper syntax, he will be displayed with a "successfully registered" dialogue on the screen.

A *password* is a secret (typically a character string) that a claimant uses to authenticate its identity. Using a password with a user identifier, such as a username, is one form of *identification and authentication*.1 *Identification* is a claimant presenting an identifier that indicates a user identity for the system. *Authentication* is the process of establishing confidence in the validity of a claimant's presented identifier. Such a password should be so strong enough so that it should not be revealed easily by anyone to avoid misuse.

A general study gives the following table which displays the number of characters that are used in a password and also the percentage of digits, uppercase and lower case characters that are used in the password.

Table 2.1: Password contents along with character length



As per our study any password that the user defines usually contains the following details as per the order of their priority

1. Lower case letters
2. Lower case letters with digits
3. Lower case letters with special symbols
4. Lower case letters with digits and special symbols
5. Upper case letters with special symbols and digits
6. Upper case letters only
7. Digits only

There is no other alternative apart from the above options for a user to define his password. Another logical aspect that our observation says is that a password defined by any engineering student (as we took a sample of engineering students for our study) contains the following interesting aspects i.e.

The password defined by them usually contains

a. Name
b. Registration number
c. Their dear ones name
d. Mobile number and
e. First name (surname)

Possibly they use only two special symbols in their password i.e.

- Underscore (_) and
- A dot symbol (.) (usually called as a period)

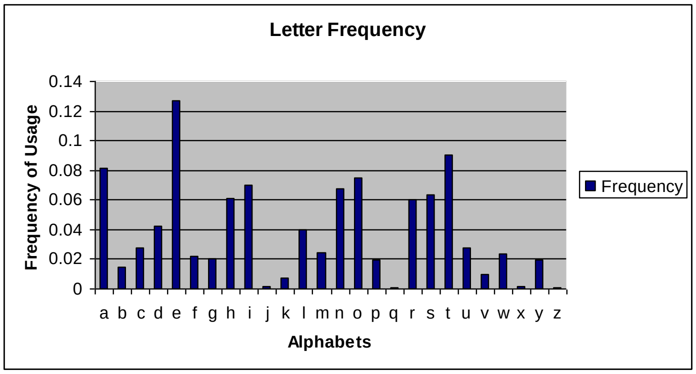
These are the password characters applied by not only the engineering students but also with all of us who perform online transactions.

III. PREVIOUS WORK

Our observations say that as the password mainly consists of characters either upper case or lower case it is easily possible for us to crack the students e-mail password. As the students try to use those passwords that are easily remembered they avoid using special symbols and digits. The number of characters are limited it is easy for us to crack the students / users passwords.

Letter frequency is a study which deals with finding the occurrence of a character either alone or along with few other characters. The following is a well known chart which displays the frequency of each character in English language [1].

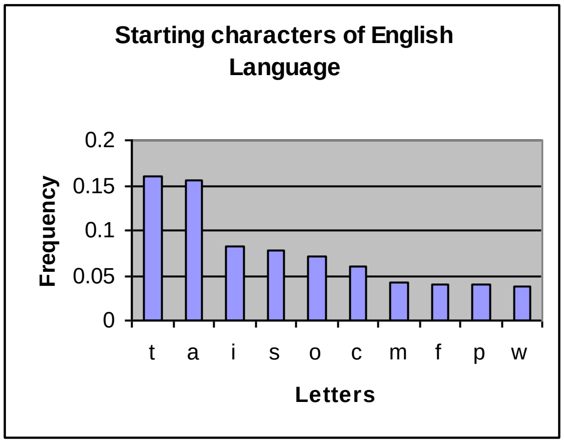
Table 3.1: Letter Frequency of English Characters



From the above chart it can be clearly observed that the character 'e' of English is used with high frequency than any other characters. Next comes t, a, o, n and so on. The least frequency is for the characters z, x, q, j and so on. So this chart says that if the user uses a character as a password either in upper case or lower case, character has the highest possibility of occurrence.

The following is a chart which displays the possibility of top ten characters of English which comes as the starting character of an English word. From the chart below it can be observed that the characters t, a are high used as starting letters of an English words. The next place is occupied by i, s, o, c, m and so on. With the above information we can come to point that the password used by the user may start with any of the starting letters which we have given below and may contain the characters as per the letter frequency which we have given in the table 2.1. On combining both the logics together we can guess the password defined by the user [2].

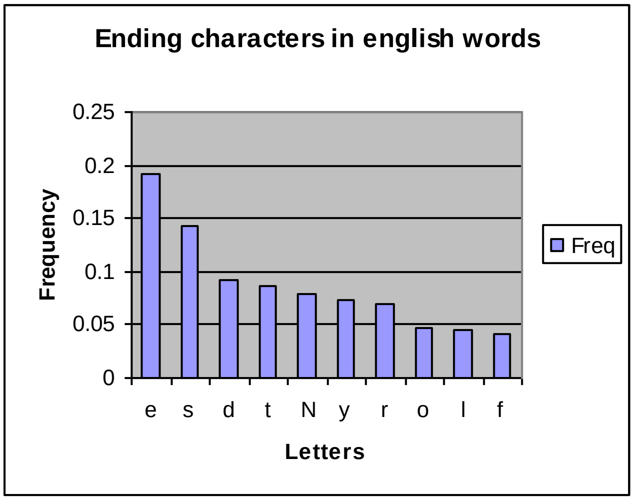
Table 3.2: Chart showing the starting letters of English language



Now as we understood how the letters are joined together in English language we step forward in understanding the next level of letter combination.

The following chart explains the frequency of characters that usually comes at the end of an English word. These characters are also helpful in predicting the password. Usually the hacker tries to join the starting and ending characters and tries place a possible character in the middle in a brute-force method. There is a possibility of knowing the password when brute-force method is applied by joining these characters [2].

Table 3.3: Chart showing the ending letters of English language



The most common digrams (in order) [2]:
th, he, in, en, nt, re, er, an, ti, es, on, at, se, nd, or, ar, al, te, co, de, to, ra, et, ed, it, sa, em, ro.
The most common trigrams (in order) [2]:
the, and, tha, ent, ing, ion, tio, for, nde, has, nce, edt, tis, oft, sth, men
The most common quadrigrams (in order) [2]:
1. that 2. ther 3. with 4. tion 5. here 6. ould 7. ight 8. have 9. hich 10. whic 11. this 12. thin 13. they 14. atio 15. ever 16. from 17. ough 18. were 19. hing 20. ment

If we use only the English words, it is very clear from our study that these words can be easily predicted and can clearly specify your password with out having any knowledge on the encryption method, algorithm followed.

IV. PREVENTION POLICIES

Now it is clearly observed that the password being a common English word can be easily revealed by just combining all the possible combinations from the above study, we have to be careful in defining ones own password which should be strong enough that should not reveal your password.

A. *Username should not replicate your behavior*

The common mistake that we usually make is we define a username such as an e-mail id which indirectly specifies your behavior. The priorities that you have followed in the username speak of your character and may easily give a clue of the password which you have defined. For example, if a user creates a email id with his college registration number then it can be easily predicted that his password may also contain either the registration number or his mobile number. For example if a student creates a username which contains his year of birth then it says he is very much concerned about himself and the password may contain any information which is completely about his details.

Jump Start

This approach of guessing the passwords by the username is proved to be true. From our studies we were able to guess nearly 12% of the passwords and 73% of our guess was found to be true in guessing the passwords.

B. *Password defining policies*

Now-a-days there is no means of security as the computer technology is increasing day by day. No organization or a bank can guarantee that their online transaction system is completely secured and reliable. When such a situation exists it is the prime duty of the user to be careful in protecting his password. It is ones own responsibility to protect his password rather than in depending on the others or a third party software.

Few methods that can be followed to avoid such type of password guessing can be listed as follows:-

i. Use long length passwords

The first solution for this problem is to use long length passwords so that the combination of words may not be predicted. Any password with a single English word can be guessed but the combination may not be estimated properly. Brute-force attacks showed good result when the password consists of a single word. But it multiple words are used then the brute-force may not work.

Example:- an email id microsfthr@gmail.com may contain a password which may include the word “Microsoft” of “hr” or some other word which is specific to the organization. On applying the brute force attack it can be easily revealed. But if the combination of such words is applied as password it may not be that much easy to crack the password.

ii. Use possible number of digits and special symbols

As the English language characters are limited (i.e. 26) the possible number of combinations we have with them is limited and can be easily predicted. But when digits are used, the possible number of combinations using the numbers is unlimited and a attacker may not be able to predict these numbers out of infinity. Special symbols can also give enough security as these symbols have a character of confusing the attacker. As these symbols don’t have any meaning the usage of them is of user’s choice. The hacker may not know the symbol used and the purpose of symbol being used.

For example :- a@d.m#i_n can be a secured password.

iii. Use mnemonics for the password

Instead of using the complete word as your password it is better to use the starting letter of a sentence of your choice. It may be easily predicted in guessing the words but the sentence, and the starting characters of the sentence can not be guessed even with brute force attacks.

Example:

Sentence	**password**
Sreekavitha engg college	SKEC
I love my country	ILMC

iv. Use shortcuts of the words

Another alternative for defining the passwords is to use shortcut of the words in a sentence. These shortcuts may also be digits and special symbols. The combination of letters, digits and passwords will be strong enough for a secured password.

Example:

Sentence	**password**
Come to me at two	cm2mi@2

v. Replacing characters with other symbols

Few characters can be replaced by the symbols or digits so that the password will be of our well known remembering password but the actual phrase that you type in the password field will be different.

Example:-

Sentence	**password**
Mymailidbank	m!m4!l!@

V. IMPROVEMENT AND FUTURE WORK

From our discussions above we clearly understood that defining a password is not that easy and care should be taken in deciding your username as the user name may also reveal your password. The expected future improvement in defining the user name and password is to use digits even for the username and the passwords. Few banks are already implementing the system of defining the username using the digits. If this is implemented in e-business, e-mails and all other online transactions then it would be much better. We expect this in nearby future.

VI. CONCLUSION

We take this is an opportunity in alerting all the users who are doing online transactions daily. We use this platform as a means of serving others and educating the net uses.

ACKNOWLEDGMENT

We express our sincere gratitude and respect to all the friends who helped us in completing this paper. Our sincere thanks go to our students who helped us a lot in surveying the required data. We are also thankful to our head of the department for his valuable suggestions in completing this paper.

REFERENCES

[1] “on letter frequency effects” by Boris New,Jonathan Graniger, Elsevier publications ,Acta Psychologica, 322-328

[2] “Case-sensitive letter and bigram frequency counts from large-scale English corpora” by MICHAEL N. JONES and D. J. K. MEWHORT, Behavior Research Methods, Instruments, & Computers-2004, 36 (3), 388-396

[3] On the Entropy and Letter Frequencies of Ternary Square-Free Words by Christoph Richard & Uwe Grimm

AUTHORS

First Author – Gantela Prabhakar, M.Tech(CSE), SreeKavitha Engineering College,Karepalli,Khammam (AP), India.
E-mail id - galexprabhu@gmail.com

Second Author – Cheldi Sarath,M.Tech(CSE), SreeKavitha Engineering College, Karepalli,Khammam (AP), India.
E-mail id - sarath_1976@rediffmail.com