Password strength and memorability

Keszthelyi András

Óbuda University, Faculty of Economics,
Institute for Organizing and Management
address: H-1081 Budapest, Népszínház u. 8., Hungary.
e-mail: Keszthelyi.Andras@kgk.uni-obuda.hu

Users can be authenticated by three different ways: on the basis of what they know (password/pin), or what they have (smartcard etc.), or what they are (biometrics). The price and complexness of these methods increases in this order. One would think that the more complex or more expensive a method the better its efficiency is. This is not true by all means. A high enough security level can be reached by the cheapest and simplest method, by applying passwords. Of course, there are a few rules of thumb must be kept in mind.

Keywords: data security, user authentication, password strength

JEL code: L-86 Information and Internet Services; Computer Software

1 Introduction

It is of critical importance to identify users in a computerized environment. User identification should be done a) in an automated manner, b) as quickly as possible, c) with the least possible error count.

The first, simplest and oldest, way, is to identify users by something they know, i.e. what only the given user can know. This is typical in case of logging into computers and network resources (password) or in case of using bank cards (pin).

This method is very simple, indeed, and because of this very simple, too. To search for a given username-password pair in a locally stored list has the difficulty of the third or fourth programming lesson.

Because of its simplicity and traditionality not only the possible attack are well known but the defence against them as well. Using passwords can be as secure as even possible, if some rules are taken into account. These rules can be decided easily by understanding how password checking possible attacks work.

Identifying users by the help of something they own can be as easy as password-checking, only the price of the equipment may be relevant. Typical examples for this method are the so-called RSA-key of our department secretaries for the Neptun, or to send the users a one-time password to their previously registered cellphones. The equipment, which possession is the basis of the user identifying, can be stolen, of course. The possible ways of stealing something are much older than the methods to get a password.

Biometrical identification methods are the newest, most complex, and most expensive and the youngest ones. So we have the least experiences in case of biometrical identification, so this is the field, where big surprises may occur. Also, biometrical methods have an additional problem: false positive and false negative errors (FAR, FRR), which means that it is normal to accept a non-authorized user (once in every 500-1.000.000 case), or to reject a valid user. FAR and FRR values are not independent, decreasing FAR will result in the increasing of FRR. There is always a finite probability of a false positive or negative result. See e.g. table 1. in (Hong, 1998., p. 1305.).

In case of the first two methods (password and possession) there is no possibility of false rejections or false acceptances.

2 How passwords work

Computer systems normally do not store the passwords themselves. Instead, they store a hash, which can be computed very easily from the password the user gives, but from the hash the real password cannot be computed. The stored hash of the real password is named as the shadow password.

Because passwords and pins are definite values and random estimating errors are impossible, they might be the best methods for user identification. Passwords, of course, can be stolen in many different ways. There is a wide range of possibilities to do that from hidden cameras to hardware keyloggers. These methods and their usage is out of the bounds normally, and not part of this paper.

Naturally, passwords can be guessed, or can be found in some systematic manners. Guessing a password can be done by trying to log in into a system using the username and applying the possible passwords. In most cases after a few (typically three) login failures the account is disabled for some seconds. This means that it is practically impossible to find a password this way, unless you can do it in a dozen or so tries. In case of bank cards you have a four-digit pin code, which would be very weak, but after your third unsuccessful trying the card will not be given back to you.

The most dangerous case if the shadow passwords are stolen in some way. In this case there is no time limit for the systematic guesses, and when the real password is found, it then may be applied in the real computer system.

We have three different ways to find the real password, if we have the shadow password.

The first way is the easiest. This is when a close logical connection exists between the password and the user's login name, between the password and the person, or the password is a frequent default one. Examples: peter – retep; peter – peter91 (let us supposed that Peter was born in 1991); username – asdfgh. Using these kinds of passwords is a serious security risk and means a serious irresponsibility of not only the users but of the system administrator and of the password policy as well.

The second method is to guess systematically on the basis of a dictionary which contains usual and frequent passwords. A program encodes them one after one, and if the two shadow passwords (the captured and the computed ones) are equal, then the real password is found. So to have such a password, which may be contained by any word-list or dictionary, is not a good choice as well.

The third and only sure method is the brute force method: if someone tries all of the possible passwords, he/she will find the one he or she needed for sure. In this case the only hope is the time factor. If the password is long and complex enough one will need too much time, maybe years, to find the appropriate one.

Because of this the most frequent advice is to use passwords like this: Axc17#4KMw. (Independently of the fact that the period is part of the password or the period at the end of the sentence). So a password is considered to be strong enough, if it has a length of 8-10 characters, containing lower and upper case letters, numbers and punctuation marks, without any meaning. The new problem is that it is hard to remember such a password. (Dinei, 2007.)

"If humans were not required to remember the password, a maximally secure password would be one with maximum entropy: it would consist of a string as long as the system allows, consisting of characters selected from all those allowed by the system, and in a manner that provides no redundancy - i.e., totally random selection." (Yan, 2000.)

"We find that a 'diminishing returns' principle applies: in the absence of an enforced password strength policy, weak passwords are common; on the other hand, as the attack goes on, the probability that a guess will succeed decreases by orders of magnitude." (Dell'Amico, 2010.)

Unfortunately, people usually do not like strong passwords in the above meaning. "A good password, in terms of the above discussion, should aim to be reasonably long, use a reasonably large character set, but still be easy to remember." (Dinei,

2007.) So there is a big difference between strong and good passwords. A good password must be strong enough, but a strong password need not to be good.

In Hungary at the end of 1999 the machines of an internet service provider (Elender) were cracked and the shadow passwords became known to the public. Security experts at the Technical University of Budapest used the shadow passwords for some examination. They tried to find the real passwords and fount a big subset of them. The statistic data of the password attributes is very instructive.

They could find 23,3% of the passwords (which means 7643 passwords out of 32796) in a few days. Some attributes of the found passwords are as follows:

only numbers	10,52%
top 25 passwords	7,31%
<=3 chars	3%
4 chars	18%
5 chars	24%
6 chars	31%
7 chars	13%
8 chars	11%

The average age of the passwords was nearly a whole year, the top 100 passwords covered 1100 user accounts. (Vajda, 2000.)

3 The cost of the brute force attack

The cost of a brute force attack may (must) be counted as the number of tries which needed to find the appropriate password in the worst case. The number of tries, of course, can calculated to the time needed to do the task supposing a given amount of computing resources.

The number of tries depends on two circumstances: the length of the password and the number of the elements of the character set it may consists of.

Let us see the number of tries in some different cases. Let us see how many different password combinations exist in case of the different character sets. Table 1 shows how many different characters there are in some different, but typical character sets. Table 2. shows how many different combinations may exist in case of the different character sets, supposing a 4-digit password length. Table 3. shows the number of combinations in case of the 26 digit character set (lower or upper case letters only) with different password length.

chars	*char set*
10	only numbers
26	only lower or upper case letters
52	lower+upper case letters
62	lower+upper case letters+numbers
72	lower+upper case letters+numbers+punctuation marks

Table 1.
Possible character sets

char set				
10	*26*	*52*	*62*	*72*
10 000	456 976	7 311 616	14 776 336	26 873 856

Table 2.
Number of combinations in case of 4 digit password length

pwd length	*combinations*
4	456 976
5	11 881 376
6	308 915 776
7	8 031 810 176
8	208 827 064 576
9	5 429 503 678 976
10	141 167 095 653 376
11	3 670 344 486 987 780
12	95 428 956 661 682 200

Table 3.
Number of combinations in case of 26 letters
and different password lengths

According to the above tables we can see that increasing the length of passwords results in a higher number of combinations than to increase the elements of the character set to be used. In other words the critical factor is the length and not the number of the character set you may/should use.

If we count with 1000 tries per sec we get that a 5 digit long password could be cracked by brute force about 1,67 minute, while a password with the same length but a 72 element charset would need a bit more than 3 weeks.

So, if the goodness of a password means that the password must be strong enough but still easy to remember, we get that the general rule, which is recommended by most system administrators and enterprise security policies, is not true and results in user passwords which cannot be remembered of, in most cases. If a password cannot be remembered of easily, user will write that password onto a piece of paper and will store it close to the computer where he/she usually have to use it.

So the final question is that how we can generate passwords of long enough but still easy to remember? One possible method is the following:

1. Choose a piece of text which you can remember of, even your favourite citation. E.g. 'She sells sea-shells'.

2. Choose some extra elements which have their own meaning for you, too, so which are easy to remember as well. E.g. 13 as your favourite number.

3. Combine the two in some way, perhaps change the statement into question etc. E.g. 'Sells she 13 sea-shells?'.

It will long enough (in our example 25 characters), random enough, which means that it is impossible to crack it by brute force (would need too many thousands of years). What do you think the easier to remember: an 8 character long random password or the example above? What if users must change their password regularly, and/or they have 10-20 password-protected accounts?

4 Teaching and learning

Investigating the skills and knowledge of the students in the fields of programming and of database management we find an alarming situation. 98,7% of the students of eighth grade have never been taught to any programming language nor any algorithms, which is worse. (Kiss, 2008.) In the homeland of John von Neumann, in the informational age.

Under these circumstances it is of critical importance to teach only methods and rules which are realistic.

Data security and data protection are in close connection with each other. To protect personal data is obligatory according to Act LXIII of 1992 on the Protection of Personal Data and the Disclosure of Information of Public Interest, especially in case of special data (personal data concerning health, addictions etc.). Understand and perform global telecommunications and tools, protection of personal data in health care are specific professional skills in health sector. (Garaj, 2010.). So it would be very important in the entrepreneurship education to the health sector to see clearly a problem and method of critical importance.

It is obvious that students will reach better results if they have not only multimedia lectures instead of traditional ones but are shown problems in connection with their own interests as well. In our age even undergraduate students have less or more (rather more) password-protected accounts. It is of critical importance to teach them the proper use of passwords, and it would fit perfectly into the field of data encryption. (Kiss, 2010)

5 Summary

It is advised in general that one must have a password of at least 8 characters long and it must contain lower case letters, upper case letters, numbers and punctuation marks as well. On the other hand there are no words about how a password can (not) be memorized. More expensive and complex methods are usually considered to be more efficient, see e.g. the usage of RSA-keys in connection with the Neptun-server of our university.

On the basis of some counting and natural logic I've shown a different method to generate strong passwords which can be remembered of easily, too. Of course, there are a few rules of thumb which must be kept in mind: (first of all) the password must be practically unable to be cracked; should be used only in a spy-proof environment; ought to be changed regularly (in case of a successful espionage).

References

[1] Dell'Amico, M. Michiardi, P.; Roudier, Y.: Password Strength: An Empirical Analysis. INFOCOM, 2010 Proceedings IEEE, pp. 1-9. http://ieeexplore.ieee.org/xpl/freeabs_all.jsp?arnumber=5461951&abstractAccess=no&userType=inst

[2] Dinei, F.; Herley, C.: A Large Scale Study of Web Password Habits. In: Proceedings of the 16th international conference on World Wide Web. ACM New York, NY, USA, 2007.

[3] Kiss, G.: A magyar informatikaoktatás vizsgálata. In: AGTEDU 2008, ISSN: 1586-846x, (pp. 163-168.)

[4] Kiss, G.: Experiences in teaching data concealment and data encryption to engineering undergraduates / 9th IEEE International Conference on Information Technology Based Higher Education and Training (ITHET 2010), Cappadokia 2010, ISBN 978-1-4244-4811-1, pp 419-423

[5] Garaj, E. (2010): Using of Moderation Techniques to Develope the Entrepreneurial Skills in Health Education. Practice and Theory in Systems

of Education, Vol. 5. Number 2. pp. 145-162, HU ISSN 1788-2591 (Online) HU ISSN 1788-2583 (Print) http://www.eduscience.hu/

[6] Hong, L.; Jain, A.: Integrating Faces and Fingerprints for Personal Identification. In: IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, vol. 20, no. 12, dec. 1998.

[7] Vajda István, Bencsáth Boldizsár, Bognár Attila: Tanulmány a napvilágra került Elender jelszavakról. BME, Elektronikus Biztonság Laboratórium, 2000.

[8] Yan, J.; Blackwell, A.; Anderson, R.; Grant A.: The memorability and security of passwords – some empirical results. Technical Report,UCAM-CL-TR-500, University of Cambridge Computer Laboratory, ISSN 1476-2986, 2000. http://www.cl.cam.ac.uk/TechReports/