Misunderstanding how Passwords Work

András Keszthelyi

Óbuda University, Keleti Faculty of Business and Management, Hungary
Keszthelyi.Andras@kgk.uni-obuda.hu

Esmeralda Kadëna

European University of Tirana, Albania
esmeralda.kadena@hotmail.com

IIt has always been an important security issue to prevent unauthorized access to our resources and it becomes more and more important as we go along in the digital era. Both possession based and biometric methods has been evolving fast and many experts say that the knowledge based methods, passwords, are not secure enough. In the present paper I prove that passwords in alone can be safe enough providing a publicly known case that shows that serious problems may occur if so-called experts give false advice to average users.

Keywords: user authentication, password

EconLit subject descriptor: L860 - Information and Internet Services; Computer Software; JEL code: L860 - Information and Internet Services; Computer Software

1 Technical background

1.1 Authentication methods

Different kind of methods to authenticate users can be classified into three groups: knowledge based, possession based and biometric methods.

First of them relies on the supposition that there is something that is known only by the legitimate user. So if someone can provide that data element, a password or PIN usually, that person should be the legitimate user.

In case of possession based methods the situation is similar. The legitimate user is the only one who possess a particular thing, a cellphone SIM or an RSA token for example, so if someone can prove that they have that particular thing they should be considered the legitimate user.

In the third case an attribute of the human characteristics, physiological or behavioural, is used to identify the legitimate user, such as a fingerprint or the dynamics of handwriting.

Each of these methods has its advantages and disadvantages. Summarizing them in short we found especially the price and the complexity of these methods can differ very much.

Passwords are for free, or at least we have no clear method to calculate its cost. Obviously, when the users don't keep the basic rules attackers may have good chances and the same goes for the lazy and not up-to-date sysadmins.

In case of possession based methods there are, at least, two different situations. First situation is when you have to give a unique object, an RFID chip for example, to each of the users, and reader devices must be applied at each entry point, as many as the required throughput would need it. Both the readers and the objects given to the users have their prices. While an RFID chip is cheap an RSA token is quite expensive not to speak about the server side software. If users use their own devices, their cellphones for example to receive an SMS for a login passcode, this method may be cheaper.

If someone decides to use a biometric method only the reader devices should be purchased, at least one device for each entry point. These devices must not be too simple unless one needs an easily hackable system and that has no points. The reader device must be smart enough to distinguish whether a living human is standing in front of it or not. If the cheap and simple face recognition system can be hacked by a printed photo of the legitimate user that system is of no use. Matsumoto Tsutomu proved in 2002 that fingerprints can be transformed into artificial gummy fingers without difficulties. [1] Nowadays a photo machine, even a mobile phone with a builtin camera, may be enough to get the fingerprint of the target person, as at the Chaos Communication Congress it was demonstrated using some photos of the German Minister of Defense from a public event [2]

Complexity is also an interesting point. The knowledge and the possession based methods need a very simple programming procedure, the data element provided for identification by the user should be checked whether it can be found in the stored list of the data elements of the legitimate users or not. This needs a simple search algorithm that is usually taught on about the third programming lesson. The result of the checking is a definite true or false depending on the circumstance whether the data element, a password for example, provided by the user could or could not be found in the stored list.

Biometric methods are rather complex, needs significantly more sophisticated analysing methods than the simple search algorithm. The more complex a system is the more possibilities an attacker may have to hack it. In addition in case of biometric methods the answer is not a simple true or false, but rather a probability. Very important quality measures of biometric methods are the false positive (FAR) and false negative (FRR) rates.

So on the basis of the outlined comparison above passwords will not soon disappear so it is our interest to see how the password based authentication can be made secure enough.

1.2 What makes a good password

It can easily be acknowledged that there are two general axiom-like requirements related to passwords. First: users should be able to keep their password(s) in mind. The second is that it must not be guessable by others.

Most users won't keep passwords like Shk17#4WLm! in mind, but more likely they will write it down somewhere and that place is within an arm's length from the computer.

The guessing methods have developed a lot since the beginning.

The most easiest for the attacker is when users use a basic password from the top ten (or top hundred) list, such as *asdfgh*, *123456*, *password*, etc.

The next level when the users selects a password that is specific for themselves, such as the date of birth, the name of their favourite actors or pets. Barack Obama's twitter account was cracked because of using this kind of a password (see below).

Another not too lucky method is when there is a formal connection between the login name as a string and the password as another string, such as *admin – admin*, *admin – admin01*, for example.

Also not a clever choice when there is an obvious logical relationship exists between the login name and the password, for example *chuck – norris*, *jamesbond – 007*, etc.

Selecting any passwords according to the above mentioned bad practices is a serious irresponsibility of the users themselves. These kind of passwords could be guessed even online. From this point onwards the online guessing won't work unless the system administrator (or the security management) is too weak The following methods usually used offline, when the shadow passwords could somehow be stolen from the system.

The next step for guessing someone else's password it the dictionary attack. In this case the attacker uses a list of probable passwords (top 100, top 1000 ones), at last

a long list of all the words from the dictionary of the given language and a program that tries out these words one by one. This means that users must not pick up any words that could be fount in any dictionaries as passwords.

As there are a lot of real life password lists came into light in the past some years attackers have the possibility to make statistical analyses to decide the most frequent password structures. The author of this paper, too, made such an analyses on the password list of rockyou.com containing about 32 million passwords that shows that passwords containing only 6-9 lower case letters were 17.3%, and six lower case letters followed by two digits were nearly 3%. [3] Knowing the most common password structures dictionary based attacks can be optimized.

The last tool of an attacker is the brute force method, when an optimized software tries all the possible character combinations – it will *find* the password, the only question is that *when*. The efficiency of the brute force attack depends on the computational capacity the attacker has and the quality of the hash function used to calculate the shadow passwords.

Publicly known test results can be found in [4,5], in 2012 Jeremy Gosney published his device and software optimized for brute force password cracking. Since then nearly four years passed and there may be organizations with stronger financial background than Gosney could afford.

As an additional result we can conclude that not the complexity of the password that matters but rather its length looking at the fact that exponential functions increase significantly quicker than polynomial functions. It can easily be calculated how much time would it need to crack successfully a password of a given length and containing characters from a given char set. Let the number of the elements in the possible char set 80. Let the length be 16, it follows that the number of all the possible character combinations are $80^{16} \sim 2.8*10^{30}$. Having a cracking speed of 10^{12} tries per sec, we would need approximately $10^{30}/10^{12} = 10^{18}$ secs, i.e. $8.93*10^{10}$ years.

Supposing that the hash function can be weak, unfortunately, I would suggest 16 characters as the minimum password length and that results in a time duration as long as the age of our universe to successfully crack the password.

1.3 Best practice at user side

Understanding the guessing methods listed above and, what is more, keeping them in mind the best practices for choosing good passwords are rather simple. Do not choose anything that is in connection with you or your login name or for what Google would give any hints. The interesting point in this that you are supposed *not* to try Google to check your password. Looking at the possibilities (and limitations) of the advanced dictionary attacks I would suggest to concatenate two or three simple words with additional modifications at least at three positions to

form a long enough password that, because of its inner logic, can be remembered of. For example, as we know that William Shakespeare was born on 26 April 1564 in Stratford-on-Avon we could prepare a password like this: Stratford26on4Avon64 with a length of 20 characters and that seems to be enough.

Additionally, the old and well-known (?) rules ought to be kept: do not tell your password to anyone else, do not write it down, do not use it on a foreign device, etc.

Usually it is advised that a user is supposed (or obliged) to change his or her password regularly. It is of no use unless a security incident has occurred or you can think of a situation in which someone would like to use your account secretly in parallel with you (the case of keeping an eye on the mailbox of the ex or the boss).

1.4 Best practice at sysadmin and/or management side

It consists of coping with three different tasks. First, the sysadmin must use all the known countermeasures to prevent online password cracking and stealing the shadow passwords. Secondly, supposing that the shadow passwords could somehow be stolen, s/he must make the brute force attack as hard as possible for the attacker, i.e. should apply a strong hash function and salting.

Thirdly, he or she should make their users choose good enough passwords, so a newly selected password should be checked if they are long enough, then against the blacklist of the top passwords and their simple derivatives.

In addition system administrators and the security management are supposed to be competent in the field of security and not to give bad advices.

2 Misbelieves

2.1 Human factor, examples for bad advices

It is hard enough to deal with the human factor in case of the average (i.e. not enough security conscious) users. What is more serious that there is a lot of obviously false advices and methods can be fount in the wild, many of them come from sources that seems to be authentic and expert.

In 2003 a guy at HP recommended a simple program [6] that could done some simple string transformations to produce strong, random-like passwords for

different sites taking the domain name and even a primitive password as input. The idea is good until an attacker does not know or suppose that it is used. If an attacker gets a clue about the usage of this program s/he could try the top ten passwords.

To use the initials of your favourite poem [7] as your password or to use the so-called leet alphabet [8] for character substituting might have been good methods until they were not advised publicly and widely.

It is advised nearly everywhere that a password should contain mixed type characters but the minimal required length does not get the necessary (if any) emphasis.

The Federal Trade Commission says "Make your password at least 10 to 12 characters long, and use a mix of letters, numbers, and special characters" [9]

Twitter also advices that "Do use use a mix of uppercase, lowercase, numbers, and symbols." [10]

According to Gmail "Include punctuation marks and/or numbers. Mix capital and lowercase letters. Include similar looking substitutions, such as the number zero for the letter 'O' or '$' for the letter 'S'." [8]

"A password based on only small letters, capital letters or numbers has a small key-space. This makes it more easy for brute-force, just because it limits the possibilities." [11]

2.2 Advice of Stanford University

Examples could be continued in an infinite series. The most problematic source I found in the near past, is the infographics on passwords at Stanford University, which is in the top ten universities in all over the world in any rankings. The first Google hint for the search expression "stanford password policy" (without the exclamation marks) is the Password Requirements Quick Guide at Stanford University IT home page. [12]

First element in their infographics: "Which characters are required in my password? – HINT: It depends on password length!" And continues that if the length is 20 characters or more it is not important which character types are included (or not included) in your password. This is correct.

In the next section they recommend that your password should be at least 16 characters in length to resist a brute force attack.

Third section: "How on Earth can I come up with a password that long?", and says "EASY! Select 4 random words." And gives the example: "orange eagle key shoe" that ends up at 21 in length (spaces included). „Now go forth and create your own awesome passwords and keep your account secure!"

For the first glance it seems to be perfect. Supposing once again the computational speed 10^{12} tries/sec and the fact that the password contains only lower case characters (and spaces) the result is: the $1.14*10^{30}$ combinations could be cracked with brute force in $3.6*10^{10}$ years, about the same as the age of our universe. Users are satisfied!

Another calculation follows, from the point of view of an attacker. Let us suppose that I would like to crack Stanford accounts. Stanford has nearly 16 thousand students and more than 2 thousand academic staff and more than 11 thousand administrative staff. Nearly 30 thousand people, at least the same amount of accounts. There are good chances that some people heed the official advice. Table 1. shows how much time would it need to crack these kind of passwords depending on the size of the vocabulary the user selects the four basic words.

vocabulary size in words	1000	2000	4000	10000
duration to crack	1 sec	16 sec	~4 min	~3 hrs

Table 1
Duration to crack passwords consisting of four basic words

So the accounts of those people would fall very soon who kept the expert-like official advise while they would feel their passwords safe.

2.3 Importance of teaching and learning

The young generation or Z-generation is considered to be familiar with ICT, especially with the use of very different internet services. Though true it may be there are some disadvantages occur as well. If these pupils don't learn the theoretical background of the technology that is part of their everyday life they will be (are) in danger. Teaching them the basics and background is especially important in the security related fields if we want our kids to avoid serious dangers. We teach them, naturally, a lot of important rules as early as we can, such as they must not obey foreign people or should look around before crossing a street. The same goes for all the internet-related staff. There are foreign people and dangerous cross-roads can be fount there, too.

“Similarly, in Israel last year, an officer from the Israeli army was arrested after allegedly forcing dozens of women and teenage girls to strip in front of their computer webcams under threat that he would attack their computers with a virus if they did not comply.” [13] This is a typical example what couldn't have happened should the victims have at least some basic IT knowledge.

Investigating the skills and knowledge of the young students here in Hungary related to the IT field, we find an alarming situation, their level is far from being

optimal. The situation is the same in Central Europe, too. [14, 15.] I, personally, think that the other parts of our world are not in a significantly better situation, especially not in the light of the above cited bullying case.

If we want to fulfill our duty to prepare our kids for a totally new era we will need give them not only technical knowledge but a new approach as well. Taking into consideration the fact, that the new generation, the Z-generation, is totally different from our generation due to the paradigm shift the widespread penetration of the internet it should be clear that we cannot teach them using only the old teaching methods and tools we have been accustomed to, we need new methods, perhaps new schools as well, as described in [16].

3 Summary

As our dependency on the day-by-day increasing amount of data stored and processed in the digital network becomes stronger and stronger the need for security is also becomes more and more important. To protect our different accounts from unauthorized access the use of passwords has been and will remain the most effective and cheapest tool. This statement is true if, and only if we use passwords correctly. To decide what is the “correct way” won't need deep professional knowledge, only a little math from our memories from the secondary school and some natural mind-set.

To stay on the safe side an important rule is not to keep any advices without some basic critics even if it comes from an otherwise highly qualified source. We can improve ourselves and/or our colleagues with training, further training and practising. It is a task free benefit in kind – even today.

References

[1] Matsumoto, T. et al. (2002). Impact of artificial "gummy" fingers on fingerprint systems, Proc. SPIE 4677, Optical Security and Counterfeit Deterrence Techniques IV, 275 (April 18, 2002); doi:10.1117/12.462719

[2] 46halbe (2014). Fingerprint Biometrics hacked again, http://www.ccc.de/en/updates/2014/ursel, 27.12.2014.

[3] Keszthelyi, A. (2013). ACTA POLYTECHNICA HUNGARICA 10:(6) pp. 99-118. (2013)

[4] Update: New 25 GPU Monster Devours Passwords In Seconds, http://securityledger.com/new-25-gpu-monster-devours-passwords-in-seconds/, 04.12.2012.

[5] New 25-GPU Monster Devours Strong Passwords In Minutes, http://it.slashdot.org/story/12/12/05/0623215/new-25-gpu-monster-devours-strong-passwords-in-minutes, 05.12.2012.

[6] Karp, A. (2003). Site-Specific Passwords, Hewlett-Packard Company, http://www.hpl.hp.com/techreports/2002/HPL-2002-39R1.pdf, 26.03.2016.

[7] Password statistics, http://computersight.com/communicationnetworks/security/password-statistics/, 16.02.2011.

[8] https://accounts.google.com/PasswordHelp, no date of uploading, downloaded 26.03.2016.

[9] Make Computer Security One of Your New Year's Resolutions, http://www.consumer.ftc.gov/blog/make-computer-security-one-your-new-years-resolutions, 26.03.2016.

[10] Keeping Your Account Secure, https://support.twitter.com/articles/76036-keeping-your-account-secure, 26.03.2016.

[11] Bakker – Jagt: GPU-based password cracking, University of Amsterdam, System and Network Engineering, 2010, p. 7.

[12] Stanford (2015). Accounts and Passwords - Password Requirements Quick Guide, Last modified December 9, 2015, downloaded 26.03.2016.

[13] Raeburn, S. (2013). Online bullying and Skype crime: Police face "new environment", The Drum, http://www.thedrum.com/news/2013/08/18/online-bullying-and-skype-crime-police-face-new-environment, 18.08.2013.

[14] Kiss, G.: Measuring Computer Science Knowledge Level of Hungarian Students specialized in Informatics with Romanian Students attending a Science Course or a Mathematics - Informatics Course / TOJET: The Turkish Online Journal of Education Technology, Volume 11, Issue 4, ISSN: 2146-7242, pp. 222-235.

[15] Kiss, G.: Comparison of the Programming Knowledge of Slovakian and Hungarian Students / Procedia of Social and Behavioral Science Journal különszám, ISSN: 1877-0428, p. 10.

[16] Karlovitz, J. (2015). Iskolarendszer-utópia vázlata In: Buda András, Kiss Endre (szerk.) Interdiszciplináris Pedagógia és az oktatási rendszer újraformálása: IX. Kiss Árpád Emlékkonferencia: Tartalmi Összefoglalók. 74 p. Konferencia helye, ideje: Debrecen, Magyarország, 2015.09.25-2015.09.26. Debrecen: Kiss Árpád Archivum Könyvtára - DE Neveléstudományok Intézete, 2015. p. 35. 1 p. (ISBN:978-963-473-841-1)