Empirical Study of Secure Password Creation Habit

Chloe Chun-Wing Lo(✉)

Sirius 16, Hong Kong, China
chloewing.lol127@gmail.com

Abstract. The general public's understanding of "secure" passwords, and how they are generated is investigated. Habits that tend to foster the creation of more secure passwords are suggested. Empirical data collected by survey participants is shown to present solid evidence that "secure" passwords created by the participants who could recall them later contained substantial substrings of simpler password chosen earlier by the participants. In contrast, those who encounter difficulty in recalling the passwords are seen to have created complex passwords substantially different from simpler ones created earlier. Some user-coping methods for the complexity-memorability dilemma are addressed, Companies are urged to adopt a salting approach before encryption, and consider new hashing mechanisms to ensure the security of user passwords. Given the limitations of human memory, it is recommended that two-factor authentication be used.

Keywords: Password selection · Password strength · Password memorization

1 Introduction

Until now, password authentication has been the major strategy to restrict access to sensitive information and services, despite the availability of alternative methods. The resistance to change seems to stem from the potential for misuse of newer alternatives [1]. Therefore, password authentication is likely to remain the most commonly used authentication method for the near future. It is essential, therefore, to help internet users to come up behaviors that increase the likelihood that the passwords they select will meet the following criteria:

1. The password must be remembered by the user
2. The password should not contain any regular or predictable pattern which makes them more susceptible to brute-force attack

The first criterion is quite obvious in its necessity. If the user cannot remember the password, the user will not be able to access the information or service. (Additionally, passwords must not be stored in a manner that makes them accessible to unauthorized persons.) Therefore, the performance of password-based authentication is subject to the limitations imposed by human memory capacity. The second requirement above heavily relies on how dictionary attack works.

D.D. Schmorrow and C.M. Fidopiastis (Eds.): AC 2016, Part II, LNAI 9744, pp. 189–197, 2016.
DOI: 10.1007/978-3-319-39952-2_19

Dictionary attack, according to Internet Security Glossary Version 2 [2], is “an attack that uses a brute-force technique of successively trying all the words in some large, exhaustive list.” In particular, using repeated substrings in passwords accelerates dictionary attack, facilitating the compromise of password security systems. The use of repeated substrings is a password security problem, since it reduces the size of the search space, and provides syntactic and semantic clues that can be exploited by an attacker.

Consequently, one way to address the password security problem is to require users to make up passwords that are as “random” as possible. In more technical terms, passwords having higher “entropy” are more less “predictable”, and so, more secure.

This, of course, increases the challenge to human memory.

In psychological terms, “High entropy” can be translated to “low associative value”. Associative value is the number of connection of a string to meaningful content made by a human being [3]. Most words one finds in a dictionary are considered to have high associative value, making them susceptible to dictionary attacks.

This sets up a dilemma between password strength and human memory: a password with higher entropy, which is more secure, has low associative value. This means those passwords are harder for human to remember [4]. Indeed, Yan et al. [5] had found empirical evidence for such intuitive phenomena, as they found that the participants who were given a randomly generated password needed to write the password down in order to memorize it.

One proposed underlying mechanism for such a trade-off in password strength and memory may lie behind the theory that human relies on phonological measures to commit information to memory. The current model of human’s working memory suggested that it contains a phonological loop and a visual scratchpad [6], to which was later added an episodic buffer [7], enslaved to a central executive which regulates the attention.

The phonological loop provides a pathway to long-term memory formation by repeatedly rehearsing the information in one’s own head. Strings of random characters, however, are often unpronounceable. This has been shown by Shallice, Warrington and McCarthy [8] to decrease their memorability.

Although this dilemma is well understood, it is not clear how the current internet users are adapting to the demand for ever-more-complex passwords. This study considers some user-coping mechanisms seen in collected data, as users’ attempt to create “more secure” passwords.

It is here initially assumed that human’s cope with the memory demand of more complex passwords by recycling and altering some simpler extant passwords (which, in any case, will not likely be “secure” either). That is, users attempt to remember their more secure passwords because they are embellished reiterations of simpler ones. It is crucial to recognize that this is not regarded by users as “reusing a password”, yet poses an equivalent risk. For, the mere existence of substrings commonly present in simple passwords leaves hints for attackers; the full reuse is not necessary.

2 Method

2.1 Participants

Participants were invited from a global community of general users to participate in a demographic survey conducted on the internet. There were in total 149 completed sets of data. The data from the participants who were not able to recall their simple passwords were culled, since this level of incompleteness did not allow the computation of the necessary statistics. 93 sets of data remained available for analysis.

Among these 93 participants, 49 were male and 44 were female, with their age ranging from 16 to 66 and older, while a vast majority falls evenly within the range of 16–45. Most of them rated their computer literacy as "moderate" or "advanced" (Table 1).

Table 1. Demographic details of participants

Variable	Category	Total	Percentage
Gender	Male	49	52.7 %
	Female	44	47.3 %
Age	16–24	33	35.5 %
	25–35	23	24.7 %
	36–45	27	29.0 %
	46-55	0	0.0 %
	56–65	7	7.5 %
	66–older	3	3.2 %
Computer literacy	Basic	6	6.5 %
	Moderate	46	49.5 %
	Advanced	32	34.4 %
	Expert	9	9.7 %
Total participants		**93**	

2.2 Measure

The complexity of the passwords was measured by the double natural log of Kaspersky time-to-break score, a widely used commercial standard. The large range of the raw scores were not convenient for comparison, so the natural logarithm were taken twice to make the values readily commensurable.

The similarity between passwords was measured using Dice's Coefficient. Given two passwords, both bigrams and trigrams were counted for the calculation of Dice's Coefficient, according to the following equation:

$$s = \frac{2n_t}{n_x + n_y}, \tag{1}$$

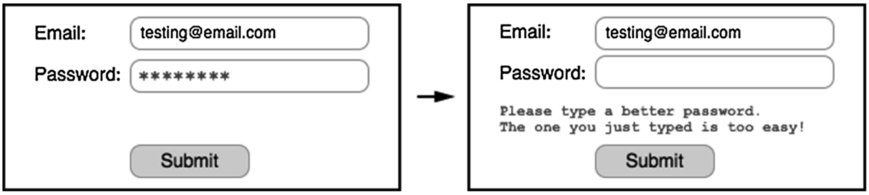
where n_t is the cardinality of the intersection of the bigrams or trigrams of the two passwords, and n_x and n_y are the cardinalities of the bigrams or trigrams of the simple password and the secure password, respectively. Dice's coefficient was chosen for its emphasizes on the existence of repeated substrings, which, as noted in the Introduction, is a major concern for password security.

2.3 Procedures

An electronic questionnaire was set up on the internet which required each participant to complete a survey. Respondents were not told that the purpose of the study was analysis of password selection strategies.

Before presentation of the survey, the participant was asked to create an account by filling in their email as the username, and providing a simple password. After the participants filled in their password, no matter how simple or secure the password was, the participants were prompted to revise their password and input one that is "more secure" (Fig. 1).

At the end of the survey, the participants were asked to recall their initial password, and their revised password (Fig. 2).

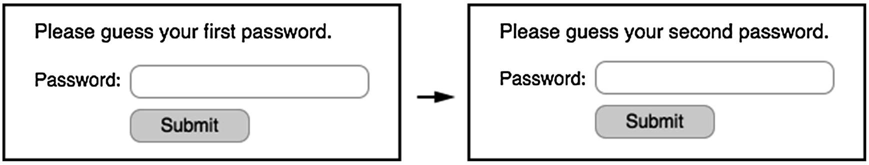


Email: testing@email.com
Password: ********

Email: testing@email.com
Password:

Please type a better password.
The one you just typed is too easy!

Fig. 1. A prompt for a more "secure" password after their first password



Please guess your first password.
Password:

Please guess your second password.
Password:

Fig. 2. A request to recall both passwords at the end of the survey

3 Result

First, a paired two-tailed t-test was conducted to compare the complexity of the simple passwords and the secure passwords. This was done to obtain an instance of a password the user believed was "more secure" than the initial one. The null hypothesis of the statistical comparison conducted was that there was no difference in the complexity of the simple password and revised password.

The result was that the revised passwords (M = 2.204, SD = 1.266) were significantly more complicated than the simple passwords (M = 0.830, SD = 1.627), $t(92) = 7.4796$, $p < 0.0001$.

An independent two-tailed t-test was conducted to find how similar the "secure" and simple passwords were between the group of participants who could successfully recall their more complicated passwords (72 of them) and those who could not (21 of them). The null hypothesis was that there was no difference in the similarity of the revised and simple passwords between these two groups of participants.

The result was that the revised passwords created by the participants who could recall them (n = 72, M = 0.324, SD = 0.389) were significantly more similar to their simple passwords, against those who failed to recall the revised passwords (n = 21, M = 0.119, SD = 0.233), t(55) = 2.9958, p = 0.0041.

The password similarity measure used Dice's Coefficient calculated from bigram counts. To measure substring reuse from the simple passwords by the "secure" passwords, a similar statistical analysis using Dice's Coefficient calculated from trigrams was also performed. A similar result was obtained: users who could recall their secure passwords (n = 72, M = 0.290, SD = 0.289) had selected "secure" passwords that were significantly more similar to their simple passwords, than users who failed to recall their revised passwords (n = 21, M = 0.090, SD = 0.198), t(65) = 3.2014, p = 0.0021.

4 Discussion

4.1 Findings

The collected data suggest that the major drawback of using passwords as the method of user authentication is that human memory capacity requires the user to choose a password that is similar to what they have been previously using. It is true that when a password gets longer, it is more difficult to crack using a brute-force approach. However, this is not the case if the attacker adopts a dictionary attack approach, especially if the attackers pre-treat the wordlist to consider the permutations of several shorter classic passwords.

This might mean that passwords users regard as "more secure" (which the data indicate are likely to contain substrings of "less secure" passwords) still carry the same security risk. Further, this problem might not be detected by the "password strength algorithm" hosted by a user's internet service. This follows from the fact that study participant's password embellishments increased the Kaspersky time-to-break score in an artificial way, leaving large substrings intact. The passwords that contain substrings of the simple passwords got a pass from the "complexity test", but the presence of repurposed substrings leaves the password authentication system susceptible to dictionary attacks.

This behavior is seen in the following examples extracted from the data:

The "secure" password in row 3 should not be any more secure than the simple one, yet the Kaspersky time-to-break score is much higher than the simple one from which it was created. This might mask the risk to dictionary attack it introduces.

4.2 Limitations

A large proportion of users decided to shift to a totally different password when prompted to enter a more secure one. There were indeed 45 out of 93 participants (48.4 %) who supplied a "secure" password having a similarity score of 0 (Dice's coefficient calculated using bigrams) with their simple password.

However, this in no way means that these passwords are any more secure than those found in Table 2. It is entirely possible (and we suspect probable) that these participants maintain a collection of different well-used passwords. This is one of the limitations of this study — other ready passwords that might be reused by participants were not known. This is left for future work, and will require more extensive and subtle data elicitation techniques.

Table 2. Example of Participants recycling their simple passwords

Simple password	Complexity score* for simple password	Secure password	Complexity score* for secure password
grapes	−1	gr@pejuice	1.839210505349484
police318	1.079918299522082	Pol318e	1.647303255736618
Delete a file	2.782182225673644	Delete a fileDelete a file	3.696182122985719
mnbvcxz	−1	mnbvcxzpoiuyt	1.566006629760012

** The complexity is calculated by the double natural log of the Kaspersky time-to-break score.*

It is emphasized that the level of similarity between the simple and "secure" passwords found during this study only constitute a lower bound. If other frequently used passwords are taken into consideration, the security risk is likely to be much higher than seen here, since internet users are known to reuse passwords across different platforms [9], and reuse usually 3 or less of them [10].

Another limitation is that Dice's Coefficient is not able to detect string patterns easily visible to humans, as shown in row 1 in Table 2. The revised password "gr@pejuice" will be deemed less similar to "grape" than "grapejuice" is due to the replaced character "a" by "@". This is a visual similarity and therefore only exists in human perception at the moment. Such similarity across passwords was not detected in this study. Another notable example will be changing "password" into "p455w0rd". Yet, this apparent randomness induced by substituting alphabets to visually similar characters does not necessarily make these password significantly more "secure".

The string morphs produced by character shape substitutions will not be present in a dictionary, but they are easily readable and pronounceable for a human being. Therefore, users might assume that such substitutions significantly strengthen security. However, this method is predictable, and does not produce combinatorial growth in the password search space. It is essentially a font variation. Attackers need only update their dictionary, rendering this strategy less effective than it appears.

The survey prompted for a more "secure" password without any explicit instruction on how the password should be made. This does not sufficiently reflect how the average internet users will react in real-life situation where the prompt for a more complex passwords usually comes with a list of criteria (e.g. at least 8 characters, include at least 1 digit, do not have 2 consecutive digits). This study provides no information on how they behave when the criteria for a strong password is difficult to meet. It is speculated that the users may respond by inputting even simpler passwords which is considered to be secure by the list of criteria but actually poses a higher security risk (e.g. "a1b2c3d4e5f6" if the prompt asks for no consecutive digits). More researches are needed to have a more accurate picture on how the users respond to the demand for a stronger password in real-life situation.

5 Recommendations

Since repeated substrings are the ultimate hint which makes gaming passwords effective, it is recommended that users implement an altered salting technique. Rather than simple salt concatenation, which does not solve the substring problem at all, users are advised to interleave the salt pattern being inserted at multiple points of the password. For example, let the salt string be $a_1a_2a_3\ldots a_n$, and the string to be protected be given by $b_1b_2b_3\ldots b_{m.}$ We may combine them and form new string $a_1b_1a_2b_2a_3b_3\ldots a_nb_nb_{n+1}b_{n+2}\ldots b_m$, encrypt this combined string and store this into the database. If a user has a username "AppleSeed" and password "macintosh", combine these two strings into "AmpapclienSteoesdh". This makes the string undetectable by dictionaries and, therefore, increases the password security while maintaining memorability.

On the industrial side, for online databases, it must be assumed that attackers will infer a ***static*** salting methodology. Therefore, enterprises should consider changing the way they combine the salt with the password, or even the hashing mechanism, depending on demographic information given by the account holder. This should be done in a way that introduces combinatorial complexity.

These approaches will reduce the unsustainable demand for increasingly complex passwords by allowing users to select simpler, more memorable passwords. Dynamic salting will also reduce the likelihood of many users having the same password, which can be used to defeat anonymization when an online database is compromised.

Completely forgoing the requirement for the selection of "secure" passwords is still not recommended. Although salting can reduce the likelihood of some attacks, it will not make the users' accounts less vulnerable to attack from people who can obtain their personal information (e.g., from social media).

If the users reuse substrings of their simple passwords when a more secure password is required in some contexts, there is increased risk that others in the user's social circle will guess the secure passwords from the other simple passwords the user is currently using. Therefore, there is still a need for users to select passwords that are not easy to guess, even by members of their intimate social circle. Consequently, education on the importance of password strength and password discretion is still necessary. It is essential that users understand that letting other people know even one of their passwords poses a risk to their other passwords. Of course, no technology can remedy misplaced trust.

Finally, it is recommended that enterprises adopt some form of two-party authentication where possible.

6 Conclusion

The effectiveness of password as a secure authentication process is not perfect, with poor password selection receiving most of the blame. The world has responded by requiring stronger passwords. We challenge the notion that asking the users to select stronger passwords solves the security problem. Password authentication relies upon human memory, and it is precisely this reliance that limits its effectiveness. How human memory works greatly restricts the level of security provided by passwords as an authentication method. This is clearly shown by how participants of this study tended to reuse substrings from their simple, less “secure” passwords to produce what they believed was a “secure” one.

There are, however, remedies to the security concern posed by weak passwords. The major drawback of weak passwords, besides being easy to guess, is that they make a database of encrypted passwords susceptible to dictionary attack. Using some altered salting technique or even variable hashing mechanisms mitigates the risk of dictionary attack.

The importance of educating the public about password strength is still of utmost importance, since none of the advanced salting techniques or hash mechanisms can fully protect passwords from being guessed by people in one’s close social circle. The public must understand that people are prone to make up very similar passwords across platforms, and that this enables people in their circles to steal their account.

It is, therefore, best to incorporate additional authentication methods that are not memory intensive alongside passwords.

Acknowledgements. I would like to thank you Mr. Monte F. Hancock for the data, and his guidance and critical review to this paper. It would not be possible for me to conduct this research without his invitation and support. There will be no words which can exactly convey the heartfelt gratitude I would like to express for his trust (hopefully not misplaced) and confidence in me.

References

1. Furnell, S.M., Papadopoulos, I., Dowland, P.: A long-term trial of alternative user authentication technologies. Inf. Manag. Comput. Secur. **12**(2), 178–190 (2004)
2. Shirey, R.W.: Internet Security Glossary, Version 2 (2007). https://tools.ietf.org/html/rfc4949
3. Noble, C.E.: An analysis of meaning. Psychol. Rev. **59**(6), 421–430 (1952)
4. McGeoch, J.A.: The influence of associative value upon the difficulty of nonsense-syllable lists. Pedagog. Semin. J. Genet. Psychol. **37**, 421–426 (1930)

5. Yan, J., et al.: Password memorability and security: empirical results. Secur. Priv. IEEE **2**(5), 25–31 (2004)
6. Baddeley, A.: Working memory. Science **255**(5044), 556–559 (1992)
7. Baddeley, A.: Working memory: looking back and looking forward. Nat. Rev. Neurosci. **4**, 829–839 (2003)
8. Shallice, T., Warringtona, E.K., Mccarthy, R.: Reading without semantics. Q. J. Exp. Psychol. **35**(1), 111–138 (1983)
9. Bryant, K., Campbell, J.: User behaviours associated with password security and management. Aust. J. Inf. Syst. **14**(1), 81–100 (2006)
10. Gaw, S., Felten, E.W.: Password management strategies for online accounts. In: Proceedings of the Second Symposium on Usable Privacy and Security. ACM (2006)